THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XVI 1964 Part 1

# IMMEDIATE MEMORY AND SIMULTANEOUS VISUAL STIMULATION

BY

H. SAMPSON

*From the Psychology Department, University of Auckland, New Zealand*

Ten right-handed, right eye-dominant aircrew trainees recalled digits and colours projected simultaneously one to each eye, under five binocular viewing conditions that precluded overlap in visual fields.

Results of this experiment favour the view that Broadbent's theory of immediate memory is a special case which, though not typical, is applicable to vision. Relevance of present findings for the problem of cerebral dominance is also discussed.

## INTRODUCTION

Broadbent's theory of immediate memory (Broadbent, 1956, 1957) emerged directly from experiments employing dichotic stimulation, and has been elaborated subsequently (Broadbent, 1958) to encompass other psychological phenomena as well. An important assumption of the theory is that a particular type of response organization in recall of auditory stimuli is representative of that to be anticipated when subjects must recall stimuli presented simultaneously in other sensory modalities. Further, this manner of recalling dichotic stimuli is regarded as an especially efficient one (Broadbent, 1956, 1957).

These assumptions have recently been questioned by Moray (1960, 1961) and Sampson and Spong (1961*a*, 1961*b*). In Moray's view (1961) a separate model for each modality may be required because of the atypical nature of the auditory-memory experiment. On the other hand, Sampson and Spong (1961*b*) suggest in effect that it might be more fruitful and parsimonious to regard the present version of the theory as a special instance of a more general formulation which is more likely to evolve from experiments on visual memory than auditory memory because of the dominant role of vision in human behaviour.

The present experiment was designed to answer two specific questions to provide a basis for deciding between these alternatives for development of the theory:

(1) Is the pattern of recall associated with dichotic stimulation demonstrable spontaneously in an analogous visual-memory experiment?

(2) If so, is this type of response pattern as efficient as other ways of organizing responses in recall?

An earlier experiment (Sampson and Spong, 1961*b*) was not successful in demonstrating this mode of recall when two types of digits were projected simultaneously, one to each eye. Therefore, in the present experiment, two quite different classes of stimuli (digits and colours) were selected, and subjects were asked to recall them when the digits and colours were projected simultaneously one to each eye, under five binocular viewing conditions which precluded overlap of visual fields.

## METHOD

This experiment was designed to investigate effects on speed and accuracy of recall of (*A*) mode of recall (i.e. whether subjects recalled all colours projected to one eye before digits projected to the other); (*B*) five different binocular viewing conditions; (*C*) two types of stimuli (digits and colours); and (*D*) the eye to which the information is projected.

*Subjects.* Ten right-handed, right-eye dominant, male aircrew trainees were used whose ages ranged from 19 to 26 years ($\bar{X}$ = 20·2 years), all of whom had normal colour vision. Handedness was determined by an individually administered questionnaire composed of 20 items suggested by Humphrey (1951) as sensitive indices of manual preference, and three additional items designed by Gillies, MacSweeney and Zangwill (1960). Eyedness was determined by the finger-aiming test.

*Apparatus.* A magazine-load automatic projector was connected to two interval timers which were in turn linked to a recording milliammeter, thus permitting control of stimulus duration and interval between successive stimuli, as well as recording of these intervals on the milliammeter by a chronograph pen attachment. Slides were "gravity-fed" into the machine from a vertically placed slide holder. The timers controlled the shutter mechanism of the projector.

A standard tape recorder was used to record the subject's verbal responses, the timing of which were recorded on the milliammeter by means of a throat microphone connected to it through an amplifier.

A binocular viewer, 3 ft. 6 in. × 6 in. × 3 in. painted matt black, with a centre division down its length that effectively bisected the visual field, and across the end a day-view screen upon which stimuli for each eye were projected. A second removable day-view screen had two 2 mm. holes punched in it 30 mm. on either side of the centre division that ran down the length of the viewer. These were the fixation points.

The visual stimuli were arabic digits from 1 to 9, and six coloured discs (red, yellow, green, blue, orange, and black which was the only non-transparent disc). The diameters of the discs equalled the height of the digits. The projected size of each stimulus was about 25 mm. × 16 mm. All stimuli were mounted on 2 in. × 2 in. glass slides in pairs.

Four different types of slides were used with digits and coloured discs placed below the point of fixation in different parts of the visual field, as follows:

(*a*) Both left- and right-eye digits and colours mounted so that the outer edge of each was projected 35 mm. temporally of the fixation point and below it. This viewing condition was designated LTRT.

(*b*) Both left- and right-eye digits and colours mounted so that the outer edge of each (in this case nearest the centre division of the viewer) was projected 30 mm. to the nasal side of the fixation point and below it. This viewing condition was designated LNRN.

(*c*) The left-eye digit or colour mounted temporally as in (*a*) and the right-eye digit or colour mounted nasally as in (*b*). This viewing condition was designated LTRN.

(*d*) The left-eye digit or colour mounted nasally as in (*b*), and the right-eye digit or colour mounted temporally as in (*a*). This viewing condition was designated LNRT.

(*e*) A fifth type was also constructed with no fixation point and in which the left and right eye digits and colours were mounted in the centre of each half of the slide. This was a *control* viewing condition.

Ten slides of each type were used. The slides were designed so that the digits 1 to 9 and the 6 colours could be assigned randomly to each eye under each viewing condition with the restriction that the same number of each be projected to right and left eyes under every viewing condition.

Previous experiment (Sampson and Spong, 1961*a*) has shown that typically under these experimental conditions, digits are recalled in pairs with those that are projected to the left eye almost invariably reported first even though experimental conditions preclude binocular scanning of pairs of digits, and timing relations are comparable to those used by Broadbent (1956). Therefore in this experiment, in order to ascertain whether a response organization similar to that obtained for audition could be elicited at all in vision, two pairs of slides were prepared for each viewing condition, in one of which all colours were projected to the left eye and digits to the right, and in the other the reverse was true. These "test slides" favoured any tendency to recall all information projected to one eye before any from the other, and as digits and colours appeared with equal frequency in each eye, any tendency to group colours and digits separately should be independent of the tendency to keep one eye separate from the other. The remaining 40 slides ("*trial slides*") did not favour any particular response tendency.

*Procedure.* In a pre-experimental period, each subject was required to superimpose the two fixation points on the day-view screen which were separated by a distance of 64 mm. This was the binocular fixation used in all the experimental trials. The subject then identified three pairs of digits—one each of types (*a*), (*b*) and (*c*) above, and each of the colours projected. Preliminary investigation suggested that reversal of type (*b*) stimuli in response was characteristically associated with this type of fixation, and so perceptual reversal of stimuli on this type of slide was used as a rough confirmation of the subject's ability to fixate in the required manner.

In the experimental session that followed immediately each subject was given fifty trials, comprising five groups on ten trials, corresponding to the five viewing conditions. Each subject received a number of digits and a number of colours on each trial, the two types of stimuli together adding up to eight, and the total number of digits and colours being equal when added together over all trials in each viewing condition. A trial consisted of four slides followed by a fifth red slide signalling the end of each trial. The digits and colours composing each trial were pre-arranged so that two identical digits or colours did not appear in any trial. Particular trials were selected at random in advance for presentation to the subject.

The subject was instructed that after every trial he was to call out all stimuli he could remember, in any order, as soon as possible after a red flash appeared on the screen that followed 0·8 sec. after the last stimulus pair on each trial. Every stimulus pair was projected for 0·3 sec., with an 0·8 sec. interval between each slide.

Before a trial, except control trials, the subject was told to superimpose the fixation points and to indicate when he had done so. Then, the stimuli were projected to each eye.

## RESULTS

(1) Table I shows the percentage of instances of recall in which subjects reported stimuli of the 10 "test slides" and of the 40 "trial slides" in a manner analogous to Broadbent's (1956) subjects on an auditory-memory task. Present subjects were ranked in these terms on the "test slides" and two equal groups (I and II) so formed

TABLE I

PERCENTAGE OF INSTANCES IN WHICH COLOURS WERE RECALLED GROUPED TOGETHER

| | *Subject* | *Test slides* (*max. possible* = 10) | *Trial slides* (*max. possible* = 40) |
|---|---|---|---|
| GROUP I | 3 | 90·0 | 97·5 |
| | 8 | 90·0 | 87·5 |
| | 4 | 80·0 | 77·5 |
| | 6 | 70·0 | 45·0 |
| | 9 | 40·0 | 20·0 |
| | Mean | 74·0 | 65·5 |
| GROUP II | 10 | 30·0 | 20·0 |
| | 2 | 30·0 | 12·5 |
| | 7 | 30·0 | 10·0 |
| | 5 | 20·0 | 12·5 |
| | 1 | 0·0 | 10·0 |
| | Mean | 22·0 | 13·0 |

were used as one classification in subsequent analyses of variance of speed and accuracy of recall of digits and colours under five viewing conditions. The mean "sinistrality index" (Humphrey, 1951) for Group I was 0·13, and for Group II, 0·09.

(2) Table I shows also that subjects who grouped colours together separately in recall on the "test slides" revealed the same response tendency on the "trial slides." The rank difference correlation between performance on test and trial slides, corrected for ties, was 0·92 which is significant at less than the 0·01 level of confidence. Moreover, in 61·4 per cent. of trials using "test slides" and 88·4 per cent. using "trial slides," digits were reported before colours by Group I. In this group the greater the tendency to report colours and digits separately in recall, the more likely it was that colours would be reported after digits. This was not of course the case for subjects in Group II, who grouped few responses together in this way (see Table I).

(3) Significantly more digits ($\bar{X} = 84·4$ per cent.) than colours ($\bar{X} = 62·7$ per cent. were recalled correctly ($p < 0·01$, Table II). However, while recalling 14·1 per cent. fewer colours than Group II, Group I still recalled digits equally as well as did Group II (Table III and A × C ($p < 0·001$) in Table II).

TABLE II

ANALYSIS OF VARIANCE OF PERCENTAGE RECALLED CORRECTLY[1]

| *Source* | *S.S.* | *d.f.* | *M.S.* | *F* |
|---|---|---|---|---|
| Recall groups (A) .. | 326·92 | 1 | 326·92 | — |
| Error (a) .. .. | 15600·35 | 8 | 1950·04 | |
| Viewing conditions (B) .. | 818·92 | 4 | 204·73 | 1·598 |
| Digits-colours (C) .. | 9947·56 | 1 | 9947·56 | 7·764** |
| Eyes (D) .. .. | 1052·03 | 1 | 1052·03 | 8·211** |
| A × B .. .. | 706·05 | 4 | 176·51 | 1·378 |
| A × C .. .. | 1842·84 | 1 | 1842·84 | 14·384*** |
| A × D .. .. | 1739·91 | 1 | 1739·91 | 13·580*** |
| B × C .. .. | 244·05 | 4 | 61·01 | — |
| B × D .. .. | 843·32 | 4 | 210·83 | 1·646 |
| C × D .. .. | 0·80 | 1 | 0·80 | — |
| B × C × D .. .. | 515·58 | 4 | 128·90 | 1·006 |
| A × C × D .. .. | 192·69 | 1 | 192·69 | 1·504 |
| A × B × D .. .. | 221·07 | 4 | 55·27 | — |
| A × B × C .. .. | 458·64 | 4 | 114·67 | — |
| Error (b) .. .. | 19986·68 | 156 | 128·12 | — |
| Total .. .. | 54497·41 | 199 | — | — |

*** Significant at < 0·001 level of confidence.
** Significant at < 0·01 level of confidence.

(4) In terms of overall performance, stimuli projected to the left eye were recalled less well than those projected to the right eye ($p < 0·01$, Table II). However, Table IV which depicts the significant A × D interaction ($p < 0·001$, Table II), shows clearly this trend was typical only of performance of subjects who alternated

---

[1] These data were transformed by the formula: $\text{Angle} = \text{Arcsin}\ \sqrt{\text{Percentage correct}}$ before analysis.

colours and digits in response (Group II), and that Group I tended to recall slightly fewer stimuli from the right.

### TABLE III

Percentage of Digits and Colours Recalled Correctly by Groups I and II

| *Group* | *Digits* | *Colours* |
|---|---|---|
| I | 86.6 | 55.7 |
| II | 82.2 | 69.8 |

### TABLE IV

Percentage of Digits and Colours Projected to Each Eye Recalled by Groups I and II

| *Group* | *Left eye* | *Right eye* |
|---|---|---|
| I | 73.5 | 71.6 |
| II | 68.2 | 83.7 |

### TABLE V

Analysis of Variance of Mean Response Onset Times

| *Source* | *S.S.* | *d.f.* | *M.S.* | *F* |
|---|---|---|---|---|
| Recall groups (A) | 9.20 | 1 | 9.20 | — |
| Error (a) | 114.53 | 8 | 14.16 | — |
| Viewing conditions (B) | 7.76 | 4 | 1.94 | 4.33** |
| Digits-colours (C) | 33.70 | 1 | 33.70 | 75.22*** |
| Eyes (D) | 4.29 | 1 | 4.29 | 9.58** |
| A × B | 1.12 | 4 | 0.28 | — |
| A × C | 12.66 | 1 | 12.66 | 28.26*** |
| A × D | 0.01 | 1 | 0.01 | — |
| B × C | 0.97 | 4 | 0.24 | — |
| B × D | 3.01 | 4 | 0.75 | 1.67 |
| C × D | 3.06 | 1 | 3.06 | 6.83** |
| B × C × D | 2.11 | 4 | 0.53 | 1.18 |
| A × C × D | 1.43 | 1 | 1.43 | 3.19 |
| A × B × D | 0.66 | 4 | 0.17 | — |
| A × C × D | 1.01 | 4 | 0.25 | — |
| Error (b) | 69.91 | 156 | 0.448 | — |
| Total | 265.43 | 199 | — | — |

** Significant at $< 0.01$ level of confidence.
*** Significant at $< 0.001$ level of confidence.

(5) Analysis of variance of mean time of onset of each response (Table V) leads to the conclusion that colours ($\bar{X} = 794$ millisec.) were recalled more slowly ($p < 0.001$, Table V) than digits ($\bar{X} = 550$ millisec.) and that both classes of stimuli

were recalled more slowly ($p < 0.01$, Table V) from the left ($\bar{X} = 723$ millisec.) than from the right ($\bar{X} = 647$ millisec.). Table VI which depicts the significant ($p < 0.001$, Table V) A × C interaction, shows that subjects who grouped colours together in response (Group I) recalled them more slowly than those who typically alternated colours and digits (Group II). Colours and digits were recalled in a similar way from both left and right, but the significant C × D interaction ($p < 0.01$, Table V) summarized in Table VII, indicates that the difference in speed of recall of colours was most marked, colours projected to the left eye being recalled most slowly of all.

TABLE VI

MEAN RESPONSE TIMES (MILLISEC.) OF GROUPS I AND II

| *Group* | *Digits* | *Colours* |
|---|---|---|
| I | 567 | 917 |
| II | 533 | 670 |

TABLE VII

MEAN RESPONSE TIMES (MILLISEC.) FOR INFORMATION PROJECTED TO LEFT AND RIGHT EYES

| *Eye* | *Digits* | *Colours* |
|---|---|---|
| Left | 580 | 867 |
| Right | 570 | 723 |

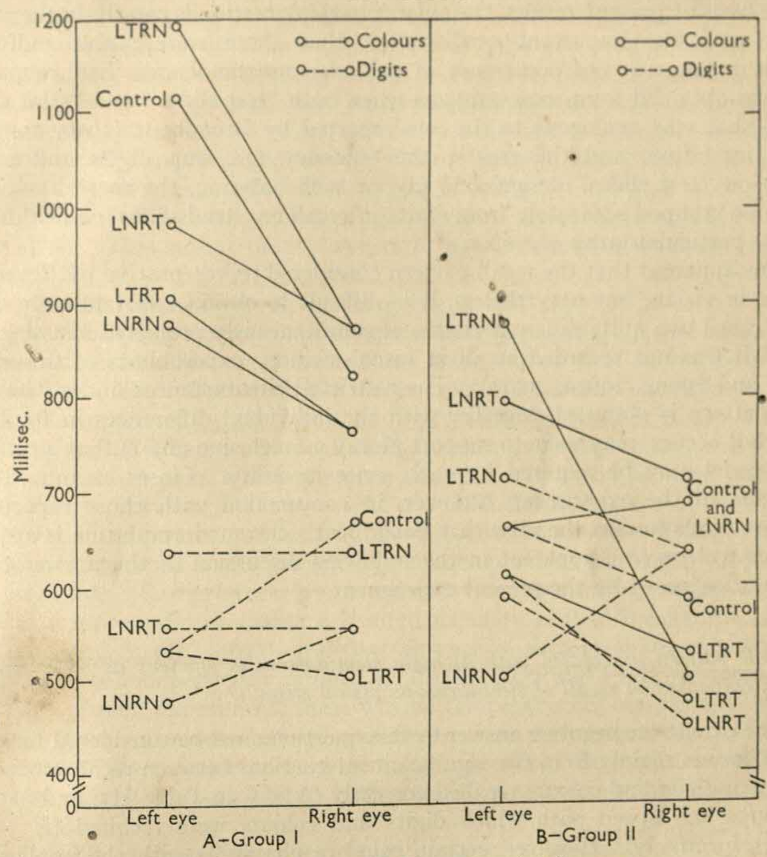
(6) There were significant differences among viewing conditions ($p < 0.01$, Table V) in speed of response onsets. Tables VI and VII, together with Figures 1A and 1B, deal with these differences in response timing in detail and reveal:

(*a*) that subjects who grouped colours and digits separately before responding (Group I) not only recalled colours more slowly than digits (Fig. 1A) but that colours projected to the left eye were recalled much more slowly still in every viewing condition than those projected to the right eye. No such left-right differences were apparent in speed of recall of digits by this same group. Indeed in both control and LNRN viewing conditions digits were recalled somewhat more slowly from the right.

(*b*) Contrasting completely with timing of responses in (*a*) was performance of subjects who typically alternated colours and digits in response (Group II). In this group, speed of recall of both digits and colours was of the same order (Fig. 1B) unlike the case for Group I, and the tendency for material on the left to be recalled slowest was general. However, digits tended to be recalled more slowly on the right in the LNRN viewing condition, just as in Group I. In addition, in Group II, recall of colours in the control and particularly the LNRN viewing conditions (Fig. 1B) tended to be somewhat slower on the right.

(7) "Organization time" (i.e. the time of onset of the first response on any trial) was independent of all experimental variables (Table VIII) and much longer than other responses ($\bar{X} = 1908$ millisec.).

FIGURE 1



Mean recall times for Experimental Groups.

TABLE VIII

Analysis of Variance of Mean Time of First Responses

| *Source* | *S.S.* | *d.f.* | *M.S.* | *F* |
|---|---|---|---|---|
| Recall groups (A) .. | 0·38 | 1 | 0·38 | — |
| Error (a) .. .. | 436·42 | 8 | 54·55 | |
| Eyes (B) .. .. | 5·16 | 1 | 5·16 | 2·52 (N.S.) |
| Viewing conditions (C) .. | 4·48 | 4 | 1·12 | — |
| A × B .. .. | 1·09 | 1 | 1·09 | — |
| A × C .. .. | 11·05 | 4 | 2·76 | 1·34 (N.S.) |
| B × C .. .. | 9·82 | 4 | 2·46 | 1·20 (N.S.) |
| A × B × C .. .. | 12·82 | 4 | 3·21 | 1·57 (N.S.) |
| Error (b) .. .. | 147·91 | 72 | 2·05 | — |
| Total .. .. | 629·13 | 99 | — | — |

## DISCUSSION

### *Is the pattern of recall associated with dichotic stimulation demonstrable spontaneously in visual memory?*

On the basis of present results, the answer to this question is clearly in the affirmative but with the important qualification that there were stable individual differences in frequency of occurrence of this response tendency. Still, a pattern of recall was obtained from some subjects when both "test slides" and "trial slides" were used that was analogous to the one reported by Broadbent (1956, 1957) for dichotic stimulation, and the greater the tendency to group digits and colours separately on "test slides" designed to favour such ordering, the more likely were colours to be grouped separately from digits in recall on "trial slides" which did not favour any particular order of response.

It seems apparent that the recall pattern considered representative by Broadbent does occur in vision, but nevertheless, it is difficult to obtain spontaneously in the sense that until two quite different classes of simultaneously projected visual stimuli were used it was not recorded at all in visual-memory experiments of this design (Sampson and Spong, 1961*a*, 1961*b*). The restricted circumstances under which the response pattern is obtained, together with the individual differences in frequency with which it occurs, may seem to support Moray's conclusion (1961) that a different memory model may be required for each sense modality. Closer examination of other results of the experiment, however, in conjunction with those reported in earlier ones, tends toward the view that Broadbent's current formulation is a special case. This will become apparent in the following discussion of the answer to the second question posed by the present experiment.

### *Is the recall pattern associated with dichotic stimulation as efficient as other ways of organizing responses in recall of simultaneous visual stimuli?*

To some extent the negative answer to this question must be considered tentative because it derives mainly from two significant interactions between recall groups and percentage of digits and colours recalled correctly (A × C in Table II); and between recall groups and speed with which digits and colours were recalled (A × C in Table V) respectively. However, certain other results agree with the implications of these interactions and so increase the likelihood that they are important. Thus, these interactions reflect the inefficiency of Group I as compared with Group II in accuracy and speed of recall of colours, and as might be expected if this were indeed the case, subjects in Group I typically reported colours *after* digits and more slowly than digits on each trial. The order in which colours were remembered by Group I suggests in turn that the slow, relatively inaccurate recall of colours may be due to a fading memory trace. It seems unlikely that this memory deficit could be an interference effect because Group II recalled colours as readily as digits although both were intermingled in response, and Group I recalled digits at least as well as did Group II.

Another reason for the difference between Groups I and II in recall of colours is indicated by results showing that information projected to the left eye was remembered less well than that projected simultaneously to the right eye. This finding is germane also to assessing the generality of Broadbent's theory of immediate memory.

In contrast with this finding, an earlier experiment, similar in design to the present one (Sampson and Spong, 1961*b*) showed that when a *different type* of digit was projected to each eye simultaneously, there were no significant differences in accuracy of recall of digits projected to either eye in a group of right-handed, right eye-dominant subjects. However, when the *same type* of digit was projected simultaneously to each eye, subjects recalled significantly more correctly of those projected to the right eye. The same trend was found in a second experiment (Sampson and Spong, 1961*a*) using similar subjects.

Accordingly, it would be reasonable to have expected present results to resemble those obtained when two types of digits were used. That is, as both colours and digits were used there should have been no differences in recall of stimuli projected simultaneously to either eye. Yet, this was not the case. Generally speaking present subjects recalled more digits correctly than colours, but overall, fewer of both classes of stimuli were recalled when projected to the left eye than when projected to the right eye. Closer inspection of these data, however, revealed an important reservation. This is that the left eye–right eye difference in recall was typical only of Group II.

A plausible inference from the foregoing is that subjects in Group I tended to form two simultaneous mnemonic organizations representing the two stimulus classes, whereas subjects in Group II formed one trace system that served both classes of information. The inference that a single trace system was operative in the latter instance is supported by the fact that in previous experiments employing only one class of stimuli, a difference in recall of information projected to each eye was typical. That is, in instances in which stimulus arrangements would *not* lead one to expect dual memory processes, a left eye–right eye difference in recall is obtained.

Timing of responses, particularly the presence or absence of *temporal grouping* of responses in recall in the present experiment provides additional support for this interpretation. Earlier work has suggested that temporal grouping of responses may be symptomatic of taxing a limited capacity central mechanism in memory (Sampson and Spong, 1961*a*), and that absence of such grouping might mean that responses were independently ordered (Halliday, Kerr and Elithorn, 1960).

In the present experiment, there was no temporal grouping of digits in Group I, although there was marked grouping of colours. In contrast, in Group II in which it is presumed only a single trace system was operative, there was marked temporal grouping of responses involving both classes of stimuli within the same speed range.

The source of these differences in recall between subjects in Groups I and II may reside in the extent to which memory processes were lateralized centrally. There is experimental evidence accumulating both in audition (Kimura, 1961*a*, 1961*b*) and vision (Sampson and Spong, 1961*a*, 1961*b*) that suggests that the left hemisphere plays an important role in perception and memory. Particularly relevant is an experiment (Sampson and Spong, 1961*b*) that indicated any deviation from being right-handed and right eye-dominant was detrimental to recall. In the light of recent work by Zangwill (1960) which shows clearly that cerebral dominance is a matter of degree, it does not seem unreasonable to postulate that although all subjects in the present experiment were right-eyed and right-handed, they may still have varied in the extent to which they were "left-brained." The suggestion then is that underlying the relative inefficiency of recall of colours by Group I was the fact that dominance was less marked in subjects in this Group and so as a consequence a division of attention occurred that permitted formation of two mnemonic processes simultaneously. This possibility accords with a recent discussion by Broadbent and Gregory (1961) of mechanisms in recall of dichotic stimuli.

It should be noted that the time required to organize a response sequence was of the same order (1·9 sec.) here as reported previously for this rehearsal time (Sampson and Spong, 1961*a*). This means that using two classes of stimuli did not affect speed of response integration.

There is one final point, though further research will be required to determine its significance. There was a tendency in Group I for digits to be recalled more slowly on the *right* in the Control and LNRN conditions, while this trend was shown clearly only for colours in the case of Group II, speed of recall under the Control condition for information from the left and right in this group being similar to that reported previously (Sampson and Spong, 1961*a*).

## REFERENCES

BROADBENT, D. E., and GREGORY, MARGARET (1961). On the recall of stimuli presented alternately to two sense organs. *Quart. J. exp. Psychol.*, **13**, 103–9.

BROADBENT, D. E. (1956). Successive responses to simultaneous stimuli. *Quart. J. exp. Psychol.*, **8**, 145–52.

BROADBENT, D. E. (1957). Immediate memory and simultaneous stimuli. *Quart. J. exp. Psychol.*, **9**, 1–11.

BROADBENT, D. E. (1958). *Perception and Communication*. London.

GILLIES, S. M., MACSWEENEY, D. A., and ZANGWILL, O. L. (1960). A note on some unusual handedness patterns. *Quart. J. exp. Psychol.*, **12**, 113–6.

HALLIDAY, A. M., KERR, M., and ELITHORN, A. (1960). Grouping of stimuli and apparent exceptions to the psychological refractory period. *Quart. J. exp. Psychol.*, **12**, 72–89.

HUMPHREY, M. E. (1951). Consistency of hand usage. *Brit. J. educ. Psychol.*, **21**, 214–25.

KIMURA, D. (1961*a*). Some effects of temporal–lobe damage on auditory perception. *Canad. J. Psychol.*, **15**, 156–65.

KIMURA, D. (1961*b*). Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psychol.*, **15**, 166–71.

MORAY, N. (1960). Broadbent's filter theory: Postulate H and the problem of switching time. *Quart. J. exp. Psychol.*, **12**, 214–20.

MORAY, N. (1961). A note on Broadbent and Gregory's paper. *Quart. J. exp. Psychol.*, **13**, 110–1.

SAMPSON, H., and SPONG, P. (1961*a*). Binocular fixation and immediate memory. *Brit. J. Psychol.*, **52**, 239–48.

SAMPSON, H., and SPONG, P. (1961*b*). Handedness, eye-dominance and immediate memory. *Quart J. exp. Psychol.*, **13**, 173–80.

ZANGWILL, O. L. (1960). *Cerebral Dominance and its Relation to Psychological Function*. Edinburgh.

*Manuscript received* 21*st May*, 1962.

# SENSORY SCALING AND THE PSYCHOPHYSICAL LAW

BY

MICHEL TREISMAN

*From the Institute of Experimental Psychology, Oxford*

Sensory scaling is analysed as a sensori-motor task for which two relations must be determined, the first, $S = f(I)$, being the relation between $I$, the stimulus intensity, and $S$, a measure of the neural effect of the stimulus, and the second being the relation between the subject's response, $R$, and $S$. It is shown that for any scaling procedure the relation between $R$ and $S$ can be arbitrarily specified, and according to the choice made $S = f(I)$ will be found to be either a power function or a logarithmic function; these two "psychophysical laws" do not reflect empirical differences between scaling procedures but conventional differences in the assumptions made when interpreting these procedures. A model is then elaborated in which $S = f(I)$ is fixed and the relation between $R$ and $S$ depends on the demands of the psychophysical task and the effects of previous learning. With this more complex model the empirical implications of the power function and log function differ; it is shown that if the latter is accepted explanations can be given for two major problems in scaling theory, the disparity between the functions given by category scaling and the direct methods, and the difference between prothetic and metathetic continua.

## INTRODUCTION

In recent years sensory scaling has attracted renewed attention, largely because of the development of new techniques, the "direct response" methods. It was hoped these would confirm Fechner's law, but this has not happened. Instead conflicts and anomalies have appeared which are not yet adequately explained. The present paper is an attempt to resolve these conflicts and to clarify the interpretation of scaling experiments.

Three findings present problems of particular interest: (1) Fechner measured sensation by cumulating "just noticeable differences" above the absolute threshold. To the extent that Weber's law held this measure was a logarithmic function of stimulus intensity. The "direct" methods include procedures such as ratio production, in which a standard stimulus is presented to the subject, and he is asked to produce or select a stimulus subjectively twice ("multiplication") or half ("fractionation") as great as the standard. From series of such judgements sensation is found to be a power function of stimulus intensity (Stevens, 1957, 1958, 1959*a*, 1959*b*, 1959*c*, 1959*d*, 1959*e*, 1960*a*, 1960*b*, 1960*c*, 1961). These two functions appear irreconcilable but no argument has been found which can finally dispose of either. (2) Methods such as category scaling and equisection are grouped by Garner (1958) among the direct methods, but are put by Stevens (1960*a*) in a special group of "partition scales." In equisection and fractionation the subject is asked to divide up a sense distance. In the latter he does so by selecting a given fraction of the sensation produced by a standard stimulus; in the former two stimuli are given and he must divide the sense distance between them into two or more equal parts. In category scaling and magnitude estimation, a "direct method," the subject assigns numbers to randomly presented stimuli. In the latter procedure he may choose his own numbers, in the former he selects them from a restricted set and he is asked to space them equally. Despite the similarities between the direct and partition methods the latter do not give a power function, but instead produce a logarithmic function, or a curve of intermediate form. (3) Differences between the "discriminability scales" (obtained by summing j.n.d.'s, or by other procedures employing the dispersion of responses to a stimulus), "partition scales," and the "magnitude scales" resulting from the direct methods are found only on certain "prothetic" continua, such as loudness or length. On other "metathetic" continua, such as pitch, all procedures give linearly related functions.

## Logarithmic Function or Power Function?

Fechner believed that a sensation could not be measured directly, but that its magnitude could be inferred if it were assumed that a constant increase in sensation magnitude accompanied each just noticeable increase in the intensity of the physical stimulus. Stevens (1957) rejects "the transforming of unreliability, inconstancy, or confusion into units of measure" and believes Fechner's assumption has been shown to be wrong because "in terms of their psychological magnitude, as measured by the scale we first set up by fractionation and magnitude estimation, the j.n.d.'s get larger as we go up the scale. Q.E.D., the j.n.d. is not constant in subjective size." As Garner (1958, 1959) points out, this begs the question by pre-supposing the validity of the power function; he argues that discriminability methods are far more reliable than fractionation and therefore it is "the discriminability criterion . . . which is directly meaningful to normal observers." Neither argument, however, is conclusive.

An intractable and inconclusive dispute may need the reconsideration of basic assumptions if it is to be resolved. Since Fechner's day there have been many advances in the understanding of measurement and scientific theory. Measurement is now understood as consisting in the assignment of numbers to objects or events according to rules, rather than as a process of discovering features inherent in them. The problem, in developing a scale of measurement, is to choose appropriate rules. More than one set of rules may be usefully applied to the same subject-matter; for example, both dynes/cm.$^2$ and decibels are valid measures of sound pressure. Neither is the "correct" scale: we choose between them according to their usefulness or the facility of the measuring operations required. If two scales do not agree, then either what they are measuring differs, or each results from the application of different rules. It is also accepted that a scientific theory is essentially a logical system which allows observations to be predicted. Only by agreeing or disagreeing with a prediction can an observation tend to confirm or refute the theory (Campbell, 1920; Stevens, 1951; Braithwaite, 1953). One implication of this is that if behaviour is to be given a scientific explanation, then any observation, even if it is a verbal statement by a subject with the form of a reference to terms in the theory, must be treated as providing evidence only through its relation to prediction. To treat it as information about the theory by virtue of its "meaning" would be as fallacious as to accept a patient's complaint of "gnawing pain" as evidence that he is prey to the champing jaws of microscopic organisms. These points will not be discussed at length here (Treisman, 1962). Their implications have sometimes been overlooked, but may be capable of resolving the power law—log law conflict. To illustrate this the summation of j.n.d.'s will be compared with ratio production.

To determine the differential threshold we might present pairs of stimuli to a subject, each consisting of a standard stimulus with a value $I'$ on a physical scale of stimulus intensity, and a variable stimulus $(I' + \triangle I')$, and ask him to say whether the latter seems noticeably greater. We would take the value of the variable stimulus for which the subject gives 50 per cent. of positive responses, or which meets some similar criterion, as the just discriminable stimulus, $I''$, and the j.n.d. would be given by $(I'' - I')$. To use such pairs of stimulus intensities $(I', I'')$ to determine a scale of sensation requires a theory relating stimulus intensity and sensation. This Fechner supplied by supposing that a sensation had a magnitude lawfully related to the intensity of the stimulus producing it, and that an act of discrimination consisted in the comparison of two sensations: only when they diverged sufficiently would the subject report a difference. The lawful relation between stimulus intensity and sensation magnitude allowed the former to be used to indicate the latter, in analogous fashion to the use, in physics, of the volume of a mass of mercury as an indirect measure of circumambient temperature, by virtue of the lawful relation between them.

If a variable, $S$, is defined, such that $S = f(I)$, and we know that

$$I''/I' = k \qquad \ldots \qquad (1)$$

where $k$ is a constant, we need more information to determine the relation between $S$ and $I$. As Fechner believed that discrimination depended on a minimum difference between sensations he postulated that

$$S'' - S' = k' \qquad \ldots \qquad (2)$$

where $S''$ and $S'$ correspond to $I''$ and $I'$, and $k'$ is a constant. It can be shown that if equations (1) and (2) hold then

$$S = k'' \log I \qquad \ldots \qquad (3)$$

if $I$ is measured in appropriate units, and $k''$ is a constant.

This equation has the form of Fechner's law. It gives $S$ as a transformation of the physical stimulus measure, $I$. Only by virtue of a theory relating stimulus intensity to the responses made does $S$ acquire an application as an indirect measure of a process intervening between stimulus and response. To determine a relation between $S$ and $I$, the empirical information summarized in equation (1), which corresponds to Weber's law, had to be supplemented by equation (2), which has no empirical content and functions only as an arbitrary formal rule allowing numbers to be assigned to "sensation magnitudes." Had a different rule been posited, such as

$$S''/S' = f \qquad \ldots \qquad (4)$$

where $f$ is a constant, this would have defined a different and equally valid relation between $S$ and $I$:

$$S = k''' I^n \qquad \ldots \qquad (5)$$

where $k'''$ is a constant, and

$$n = \log f / \log k \qquad \ldots \qquad (6)$$

Equation (5) would then be as valid a rule for assigning scale values to "sensations" as equation (3).

We could derive a ratio production scale in similar fashion. Pairs of stimuli, $(I', I' + \triangle I')$, might be presented to the subject, and the value of the variable stimulus satisfying a 50 per cent. response criterion would be identified as $I''$. Here a question such as: "Does the variable stimulus seem more than twice as great as the standard?" would be asked. Pairs of stimuli $(I', I'')$ would then be available for constructing a "sensory scale." In this method equation (4) is always assumed to hold, with $f$ equal to the fraction or multiple ("half as great," "twice as loud," etc.) asked for in the experimental instructions, and $S$ is found to be related to $I$ by a power function like equation (5).

Just as equations (1) and (4) taken together determine equation (5), so do equations (5) and (4) determine equation (1) (Ekman, 1958). Thus for the raw data of the ratio production method, as well as for j.n.d.'s, the relation between $I''$ and $I'$ is given by Weber's law, so that a difference in the underlying processes does not account for the different psychophysical laws derived from the two procedures. The derivation of a logarithmic function from discriminability data and a power function from ratio production is wholly due to the application of equation (2) in the first case and equation (4) in the second. But since these formal equations are no more than arbitrary definitions or rules they could legitimately be interchanged; if this were done discriminability data would give a power function, and fractionation a logarithmic psychophysical law.

Since all the empirical information used is contained in equation (1), and the (sometimes implicit) assumptions made to allow a scale to be derived are purely formal, the logarithmic and power function measures have the same empirical implications here. This can be illustrated by the analogy of an "automatic photometer" which selects series of lights of graduated brightness. If we examined such a series and found brightnesses of 1, 10 and 100 millilamberts, in that order, we might suppose that the photometer measured in millilamberts and selected lights separated by a constant brightness ratio (10/1), or that it measured log.millilamberts and selected lights at a constant interval (log 10). But no possible examination of the photometer would allow us to decide between these two accounts; the same set of intervening operations could be described in either way.

Although proponents of the power function consider that "sensation has ceased to be regarded as a private mentalistic entity . . . and has joined the class of other natural constructs, defined in terms of the operations used to denote and measure it" (Stevens, 1961), they would presumably deny that the choice of equation (4) for treating ratio production data is arbitrary, and argue that it is in some sense "correct," perhaps because subjects are reporting their experiences "directly." This appears to be implied by the argument justifying equation (4): "the response of an observer who says 'this is half as great as that' is one which, for the purpose of erecting a subjective scale, can probably be accepted at its face value" (Stevens and Davis, 1938). But even this claim would not make the choice of equation (4) more than arbitrary. Since no scale can exist prior to the establishment of rules defining one, the subject cannot be directly acquainted with an inherent measure of sensation, nor can his performance of operations such as "halving" reveal such implicit knowledge. Moreover the subject's belief or statement that he is "halving" rather than subtracting is not in itself information about the proper constitution of a model or theory of the processes underlying his performance; his statement can constitute evidence about it only by agreeing or disagreeing with some statement that the model might predict he would make.

If we had begun by observing the behaviour of a subject making psychophysical judgements we might have asked "What processes underlie this?" and, since we know that a stimulus produces various neural effects, we might have tried to answer by applying a simple model in which a neural effect of the stimulus varies in some way with change in stimulus intensity, and the magnitude of this central neural effect determines the response—"Yes" or "No," "More than twice as great" or "Less than twice"—when two stimuli are presented. If we had then wanted some indirect measure of the magnitude of this intervening neural effect, we might have used for this purpose the empirical relations found between $I''$ and $I'$. By the same reasoning as before this would have led to equation (3) or equation (5), with $S$ now an inferred measure of the response-determining central neural effect of the stimulus. Since we know that neural processes intervene between stimulus and response, it is more parsimonious to make $S$ refer only to these than to give it, as well, undefined non-neural associations. If we are only justified in giving it this specific reference then the varied associations of "sensation" and similar terms become an embarrassment, and we might do better not to use them at all in discussing the function of the inferred measure, $S$. It is clear that here we would not be tempted to accept the subject's beliefs about what he is doing as information about the intimate nature of the underlying mechanisms postulated by the model. We see, too, that there is no good reason for considering ratio production a more "direct" scaling procedure than cumulating j.n.d.'s. The experimental instructions may refer to "sensation" more directly in the former procedure, but the logical basis for deriving a measure, $S$, and our model of the processes involved, appear of equal complexity in each case.

Both for Weber's law and in determining a power function from the direct methods it has been found that for low values of $I$ the best fit to data is given by adding a constant, $a$, to $I$, so that the power law could be written as

$$S_p = k'''(I + a)^n \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (7)$$

and Fechner's law as

$$S_f = n \log (I + a) \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (8)$$

(Gregory, 1956; Ekman, 1958, 1959; Stevens, 1957; Stevens and Stevens, 1960; Galanter and Messick, 1961). However $a$ will be ignored in further discussion. "$S_f$," and "$S_p$," will be used to distinguish the measures of the central effect of the stimulus given by the logarithmic and power functions.

It might be objected that although the log and power functions have been shown to be equally valid alternatives for ratio production, this would not be true for magnitude estimation, because here no assumption is made. In this procedure stimuli are presented in a random series, and the subject assigns numbers to them. The median responses are simply plotted against the stimulus values and a power function is obtained with an exponent similar to that given by ratio production. However it is wrong to suppose that this procedure gives a power function without depending on any assumption similar to equations (2) or (4).

In the simple model discussed above the processes intervening between stimulus and response are divided into two stages. First, the stimulus causes a neural effect dependent on its intensity. Then the central effects of two stimuli are compared, the divergence between them determining which response is selected. For both threshold discrimination and ratio production similar operations could be employed at this stage, the main difference being the degree of divergence necessary to produce a positive response. In this model the subject acts as a computer of sorts, an initial stage of measurement being followed by a subsequent computation. We can suppose that the central effect of the stimulus, which serves as the "computer's" measure of $I$ in its further calculation, is related to the stimulus intensity in the same way whatever the psychophysical task, and that it is the further computations which will vary to meet the demands of different tasks. (The "sensory scale," $S_f$ or $S_p$, is then an indirect measure, derived from the stimulus-response relations, of this central effect.) In magnitude estimation a different operation would be applied to the initial measure, this being its transformation into a form which specifies the appropriate response. The relation between the number, $R$, which constitutes the subject's response, and the stimulus intensity, $I$, which (ignoring constants whose value depends only on the choice of unit) is given by

$$R = I^n \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (9)$$

is then an empirical equation summarizing the experimental results, equivalent to equation (1) in this respect, and, if this model is accepted, it is wrong to mistake it for the "psychophysical scale." We can suppose it to depend on two unspecified functions, $S = f(I)$, and $R = g(S)$. Taken together, these have the same relation to equation (9) as equations (2) and (3), or (4) and (5), have to equation (1). If one of the pair is arbitrarily specified this will determine the form of the other. Thus the conventional belief that magnitude estimation shows that

$$S = I^n \tag{10}$$

is based on the implicit assumption that

$$R = S \tag{11}$$

If it were assumed instead that

$$R = e^S \tag{12}$$

then the psychophysical law would have the form

$$S = n \log I \tag{13}$$

In equations (10) and (13) we have again rules for assigning a measure to the central effect of the stimulus, and equations (10) and (11), or (12) and (13), are alternative descriptions for the same set of underlying operations.

It has been argued that cross-modality matching experiments, in which subjects are asked to equate the apparent strengths of stimuli on two dimensions, such as brightness and loudness, have "closed an important circle in a process of validation" of the power law (Stevens, 1959*a*). The argument is that if, given an appropriate choice of unit, the laws for the two dimensions can be written as $S_{p1} = I_1^n$ and $S_{p2} = I_2^m$ and corresponding values of $S_{p1}$ and $S_{p2}$ are paired, then the resulting function can be written as $I_1^n = I_2^m$. Taking logs this gives

$$\log I_1 = \frac{m}{n} \log I_2 \tag{14}$$

so that a log–log plot of the paired stimulus strengths would be a straight line with a slope equal to the ratio of the two exponents. The confirmation of this prediction has been held to strengthen the power law (Stevens, 1957, 1959*a*, 1959*b*; Stevens, Mack and Stevens, 1960).

However if the scales used for each dimension were $S_{f_1} = n \log I_1$, and $S_{f_2} = m \log I_2$, equation (14) would follow equally well, so that though confirmation of the prediction shows that subjects perform related tasks consistently it does not provide a reason for preferring equation (4) to equation (2) in defining a central scale.

Although $S_p$ and $S_f$ are based on alternative arbitrary definitions and have the same empirical implications, this does not mean that we can have no good reasons for preferring one to the other. If we compare the subject to a computer then "*S*" refers to the measure of the stimulus used by this computer, and pairs of equations such as (2) and (3), or (4) and (5), give alternative descriptions of the successive stages of measurement and calculation performed by it. In this simple model both descriptions make the same predictions, and the choice between them can be dictated only by convention or convenience. But the model can be elaborated by the addition of further assumptions in such a way that the empirical implications of the power and log functions will differ. Such an assumption, which we will make, is that error or failures are larger or more frequent when the calculations the computer must perform are complex than when they are simple, and that its mode of operation is such that calculations which appear simple to us, such as addition, or complex, such as the extraction of roots, will require it to perform in a correspondingly simple or complex way. For some tasks we would expect the computations required to be simple if the initial measure of the stimulus is given by a log function, and complex if it is given by a power function, or vice versa. Depending on which function we assume we now have differing expectations with which to compare subjects' behaviour; if it is efficient when a logarithmic measurement function would imply simple further calculations, and faulty when they would be elaborate, and this congruence between our description of what the computer is doing and its efficiency is not shown when the power function is assumed, we would have good reason to prefer the log function model.

We can now apply this more complex model to some further problems in scaling theory, in order to discover whether performance is in fact variable and prone to error when the model described in terms of $S_f$, or $S_p$, implies difficult and involved operations, and ready and reliable when simple operations are implied; whether either model needs fewer and simpler operations to provide accounts of the performance observed; and whether the models lead to explanations which are general, or arbitrary and *ad hoc*.

### *"Partition scales" and "magnitude scales"*

If the disagreement between partition and magnitude scales is to be understood a different problem must first be considered: Why do the power laws have the exponents found? One suggestion is that the power functions are the "dynamic input–output operating characteristics of the sensory transducers," i.e. the sense organs (Stevens, 1960*a*, 1961). But even if there were evidence that the afferent neural input was related to stimulus intensity by power functions with the exponents found, it would still need to be explained why the numbers chosen by the subjects for their responses should show the same relation. If measurement is a process of assigning numbers according to rules, then a subject can only make judgements that imply measures of magnitude consistently if he is following some rule, and subjects will agree with one another only to the extent that they are following similar rules. If agreement is sufficiently general to allow scales to be derived from procedures such as magnitude estimation, then some common process of tutelage must be responsible for this uniformity.

Warren (1958) has revived the view that experience with stimulus intensities is the source of this learning. He suggests that judgements of sensory magnitudes are based on experience of the correlation between sensory excitation and "the amount of some physical attribute associated with the stimulus," e.g. the association between the loudness of a sound, or the illumination of a surface, and the distance of the light or sound source. It is plausible to suppose that the ability to use numbers or measures in dealing with stimuli arises from the young child's need to predict environmental changes and anticipate the effects of his actions. Where attributes of stimuli, such as loudness and distance, vary in concert, he must learn the relation between them. Dimensions which are frequently subject to additive operations, such as visual length, area, or weight, require that he must learn to predict the effects of these. If his initial measure of the stimulus, on such dimensions, is a nonlinear function of $I$, then he will do best to transform it into a linear function before performing analogous additive operations on it in order to derive predictions. Thus if his initial measure of length corresponded to $S_f = log\ I$, he would learn to transform this by the function $R = e^S$ to give a final relationship of the form $R = I$. It is likely that sensory judgements are often determined by correlated changes in more than one dimension, so that he might transform his initial measures of loudness and brightness, for instance, into forms simply related to his measure of distance. If the final relation made loudness or brightness halve as distance doubled, then these dimensions would give $R = I^{0.5}$.

This view is supported by some of the experimental values found for exponents. Thus visual length, visual area, temporal duration (Stevens, 1957), and, when density is kept constant (Warren and Warren, 1956) or cues to the size of the weights are eliminated (Baker and Dudek, 1955), heaviness, all have exponents approximating unity. For loudness and brightness values ranging between 0·3 and 0·5 are obtained (Stevens, 1955, 1957; Warren and Warren, 1958; Warren, Sersen and Pores, 1958; Warren and Poulton, 1960). Since changes in the taste of sapid substances must often be due to their dilution a linear function of concentration might be expected: exponents found for taste range between 0·8 and 1·3 (Warren, 1958; Stevens, 1960*a*). The concentration of an odorous substance diffusing freely should obey an inverse square law: the exponent for olfaction is close to 0·5 (Stevens, 1960*a*, 1960*b*; Reese and Stevens, 1960).

Against the "learning hypothesis" Stevens argues that some dimensions, such as electric shock, are unfamiliar, and that it does not account for some of the exponents found. Thus he finds that the exponent for loudness is 0·3 (Stevens, 1955) and not 0·5 as claimed by Warren, Sersen and Pores (1958). However the direct methods are readily affected by biases arising from details of procedure (Garner, 1958; Warren, 1958); for example, when a brightness scale was determined by magnitude estimation using a 5° target it gave an exponent of 0·36, but when a point source was used the exponent rose to 0·47 (Stevens and Galanter, 1957).

We must now consider the problem of the disparity found between "partition scales" and "magnitude scales" on prothetic continua. Category scaling is similar to magnitude estimation save that the set of numbers which can be used is restricted, and they are supposed to be equally spaced. Equisection bears a similarly close resemblance to fractionation. But despite the similarities the "partition methods" produce logarithmic functions, or functions of intermediate form, but not the power functions given by the direct methods (Stevens and Galanter, 1957; Torgerson, 1960; Galanter and Messick, 1961).

This finding is embarrassing for proponents of the power function, who suppose that in category scaling the subject is attempting to divide the $S_p$ scale into equal intervals. It leads Stevens (1960*a*) to the conclusion that "observers are so constituted that they are unable to partition a prothetic continuum without a systematic bias." He attributes this bias to variation in the subject's sensitivity to differences: "Near the lower end of the scale where discrimination is good the categories tend to be narrow . . . . Near the upper end, where a given stimulus difference is less easy to detect, the categories broaden." Two related reasons are given for this: the subject fails to reproduce the power function "for the basic reason that his ability to tell one magnitude from another varies over the scale and affects the width of his categories. Since he can easily tell 0·5 sec. from 1·0 sec. he tends to put them in different categories; since he can only with difficulty tell 3·5 sec. from 4·0 sec. he tends to put them in the same category" (Stevens, 1957), or because "a given difference at the low end of the scale is more noticeable or impressive than the same difference at the high end of the scale" (Stevens, 1959*a*).

There are difficulties with each version of this explanation. No reason is given why the greater impressiveness of the lower end of the scale should not affect the direct methods in the same way as the partition methods. If their "impressiveness" leads to weak stimuli being rated too large why is their magnitude not over-estimated? If it causes the subject to bisect the distance from 0·1 to 6 sones at 0·8 sones why does it not prevent him halving 6 sones at 3 sones? The argument about the confusion of neighbouring stimuli depends on their being linearly spaced in the example; this corresponds to closer spacing, on a discrimination (logarithmic) scale, for strong stimuli than for weak. But if the discrepancy between the category scale and power function is a consequence of stimulus spacing, then varying the spacing should affect the discrepancy. The more the weak stimuli are separated and the strong stimuli bunched together the greater it should be, but if the stimuli are equally spaced on a discrimination scale it should disappear. But Stevens and Galanter (1957) found that the disparity was most marked when the stimuli were logarithmically spaced, and that it was reduced by bunching the strong stimuli.

An experiment by Torgerson (1960) further illustrates the difficulty of this explanation. Using the methods of category scaling and magnitude estimation, his subjects scaled a series of greys both for "lightness" and for "darkness." Treating the data in the conventional way, magnitude estimation gave a power function and category scaling an approximately logarithmic function in each case. Since the category scale has the same relation to the magnitude scale in each case, Stevens' explanation would imply, as Torgerson points out, that the lower end of the scale is more impressive and shows better discrimination when "lightness" is being judged, and the upper end when the subject is scaling "darkness."

A different explanation is suggested if the log function model is taken with the "sensory learning" hypothesis (Treisman, 1960). In magnitude estimation the subject's responses, $R$, will be related to $I$ by the power function resulting from his sensory experience. To produce these responses he must exponentially transform $S_f$. Though subjects may find this procedure difficult, and some cannot do it at all (Stevens, 1956), the partition procedures present an even more complex requirement. In these the subject must select the series of decreasing intervals on $S_f$ which will, when exponentially transformed, give equal intervals in terms of the required power function of $I$. If this task is too difficult for the subject he might select equal intervals on $S_f$ instead, or he might fail to make the intervals decrease sufficiently. If mean category assignments are plotted against stimuli the former case would give a logarithmic function, and the latter a function of intermediate form, such as is often found.

Some support for this explanation is provided by a recent experiment in which loudness was scaled by magnitude estimation and category scaling and the usual nonlinear relation between the two functions was found (Galanter and Messick, 1961). When the category judgements were scaled by the method of successive intervals the experimenters obtained a function logarithmically related to the magnitude scale, and they found that the category widths determined by this method "tend to decrease as scale values increase with a slight reversal at the upper extreme."

This explanation depends on the assumption that operations that appear complex in the model should be more difficult for the subject. We can translate it into terms of a power function model, but then the simpler descriptions will no longer apply to the behaviour the subject finds easier or chooses to perform. We would have to say that choosing equal intervals on the measurement function $S_p$, with a subsequent linear transformation to $R$, is too difficult for subjects to achieve. Instead they must follow the "easier" course of choosing equal ratios, or compromise between the two. This explanation appears implausible. It is unlikely that it would be derived or accepted if the power function model were adopted. This suggests that the log function model is likely to prove more useful in providing explanations and making predictions.

## THE PROTHETIC-METATHETIC DISTINCTION

For a dimension such as loudness three different functions are given by discriminability scaling, category scaling, and the direct methods. But this is not true for pitch: here all three types of procedure give the same function. Scales of the former sort are called "prothetic," and the latter "metathetic" (Stevens, 1957, 1959*a*, 1961). The explanation given for this is that "it seems evident that the existence of two basic kinds of physiological mechanisms is reflected in the behavior of the psychological scales" (Stevens, 1960*b*), these being "additive" and "substitutive" neural processes. Neural coding is probably more complex than this suggests (Wever, 1949; Licklider, 1959), but even if a consistent physiological difference between the two types of dimension were found this would not in itself explain the different relation between the procedures applied and the scales derived in the two cases.

It was suggested that the need to relate sensory information from different sources would make subjects learn to transform their initial measure of loudness, $S_f$, into a final measure, $R$, linearly related to the distance of the sound-source. But dimensions not simply and reliably related to an important physical attribute, nor subject to additive operations, would not be acted on by environmental pressures tending to reward some particular transformation of $S_f$, and so, we may assume, in these cases none would be learnt. Asked to estimate magnitudes on such a dimension the subject would use the simplest transformation of $S_f$, $R = S_f$, to determine his responses, and for category scaling he would use equal intervals on $S_f$. The scales derived by both these methods would, therefore, coincide with that given by cumulating j.n.d.'s. Pitch might be expected to be a dimension of this sort, as also dimensions for which there is no simple corresponding physical stimulus. This last expectation receives some confirmation from an experiment in which the aesthetic value of specimens of handwriting was scaled by a procedure using ratio estimations, and by the method of paired comparisons, and a linear relation was found between the resulting "direct" and "discriminability" scales (Ekman and Kuennapas, 1960).

This explanation could be expressed in terms of $S_p$. If equation (4) is applied to j.n.d.'s a power function should be given, to the extent that Weber's law holds. But on the metathetic continua conventional scaling of fractionation and magnitude estimation data gives not this curve but, instead, the cumulated j.n.d. function. This would imply that in magnitude estimation $R$ is a logarithmic function of $S_p$, and we would have to describe this as the "simple" relation occurring in the absence of environmental pressure. This is far less plausible than when $S_f$ is used as the basis of explanation.

## CONCLUSION

A number of problems in scaling theory have been reviewed. A simple model of the psychophysical judgement process was presented, the subject's performance being likened to that of a computer in which an initial stage of measurement, followed by a subsequent calculation, determine a reliable relation between input and output. The observed relation between the responses and the stimuli can be expressed as $S = f(I)$, $R = g(S)$, where $S$ is an inferred measure of the central effect of the stimulus, and the two functions can be treated as descriptions of the successive stages of measurement and computation. In terms of this simple model the conflict between the log and power functions is not a matter of substance but of definition; any scaling procedure which appears to give one could equally well be interpreted as giving the other. A more elaborate model was then developed in which alternative versions, which differed in their empirical implications, could be derived by making $S$ a logarithmic or power function of $I$. Taken with the "sensory learning hypothesis" this model suggested explanations for the disparity between the results of the direct and partition procedures, and for the difference between prothetic and metathetic continua; these seemed plausible when expressed in terms of the log function but arbitrary and implausible if the power function were used. This suggests that the log function model may prove the more useful.

## REFERENCES

BAKER, K. E., and DUDEK, F. J. (1955). Weight scales from ratio judgments and comparisons of existent weight scales. *J. exp. Psychol.*, **50**, 293–308.

BRAITHWAITE, R. B. (1953). *Scientific Explanation.* London.

CAMPBELL, N. R. (1920). *Physics: The Elements*. London.

EKMAN, G. (1958). Two generalized ratio scaling methods. *J. Psychol.*, **45**, 287–95.

EKMAN, G. (1959). Weber's law and related functions. *J. Psychol.*, **47**, 343–52.

EKMAN, G., and KUENNAPAS, T. (1960). Note on direct and indirect scaling methods. *Psychol. Rep.*, **6**, 174.

GALANTER, E., and MESSICK, S. (1961). The relation between category and magnitude scales of loudness. *Psychol. Rev.*, **68**, 363–72.

GARNER, W. R. (1958). Advantages of the discriminability criterion for a loudness scale. *J. acoust. soc. Amer.*, **30**, 1005–12.

GARNER, W. R. (1959). An argument for the use of discriminability scaling procedures in scaling sensory intensities. *Acta Psychol.*, **15**, 94–7.

GREGORY, R. L. (1956). An experimental treatment of vision as an information source and noisy channel. In CHERRY, E. C. (Ed.) *Information Theory*. London. pp. 287–99.

LICKLIDER, J. C. R. (1959). Three auditory theories. In KOCH, S. (Ed.) *Psychology: A Study of a science*. Study I. Conceptual and Systematic. Vol. 1. New York. pp. 41–144.

REESE, T. S., and STEVENS, S. S. (1960). Subjective intensity of coffee odor. *Amer. J. Psychol.*, **73**, 424–8.

STEVENS, J. C., MACK, J. D., and STEVENS, S. S. (1960). Growth of sensation on seven continua as measured by force of handgrip. *J. exp. Psychol.*, **59**, 60–7.

STEVENS, J. C., and STEVENS, S. S. (1960). Warmth and cold: dynamics of sensory intensity. *J. exp. Psychol.*, **60**, 183–92.

STEVENS, S. S. (1951). Mathematics, measurement, and psychophysics. In STEVENS, S. S. (Ed.): *Handbook of Experimental Psychology*. New York. pp. 1–49.

STEVENS, S. S. (1955). The measurement of loudness. *J. acoust. Soc. Amer.*, **27**, 815–29.

STEVENS, S. S. (1956). The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, **69**, 1–25.

STEVENS, S. S. (1957). On the psychophysical law. *Psychol. Rev.*, **64**, 153–81.

STEVENS, S. S. (1958). Problems and methods of psychophysics. *Psychol. Bull.*, **55**, 177–96.

STEVENS, S. S. (1959*a*). On the validity of the loudness scale. *J. acoust. Soc. Amer.*, **31**, 995–1003.

STEVENS, S. S. (1959*b*). Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.*, **57**, 201–9.

STEVENS, S. S. (1959*c*). Tactile vibration: dynamics of sensory intensity. *J. exp. Psychol.*, **57**, 210–8.

STEVENS, S. S. (1959*d*). On the measurement of sensation. *Acta Psychol.*, **15**, 91–4.

STEVENS, S. S. (1959*e*). Measurement, psychophysics, and utility. In CHURCHMAN, C. W. and RATOOSH, P. (Eds.) *Measurement: Definitions and Theories*. New York. pp. 18–63.

STEVENS, S. S. (1960*a*). Ratio scales, partition scales and confusion scales. In GULLIKSEN, H., and MESSICK, S. (Eds.) *Psychological Scaling: Theory and Applications*. New York. pp. 49–66.

STEVENS, S. S. (1960*b*). The psychophysics of sensory function. *Amer. Scientist*, **48**, 226–53.

STEVENS, S. S. (1960*c*). On the new psychophysics. *Scand. J. Psychol.*, **1**, 27–35.

STEVENS, S. S. (1961). The psychophysics of sensory function. In ROSENBLITH, W. A. (Ed.): *Sensory Communication*. New York. pp. 1–33.

STEVENS, S. S., and DAVIS, H. (1938). *Hearing. Its Psychology and Physiology*. New York.

STEVENS, S. S., and GALANTER, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, **54**, 377–411.

TORGERSON, W. S. (1960). Quantitative judgment scales. In GULLIKSEN, H., and MESSICK, S. (Eds.): *Psychological Scaling: Theory and Applications*. New York. pp. 21–31.

TREISMAN, M. (1960). The nature of the psychophysical law and its significance for the scaling of sensory magnitudes. Paper read at the XVI International Congress of Psychology, Bonn.

TREISMAN, M. (1962). Psychological explanation: The "Private Data" hypothesis. *Brit. J. Phil. Sci.*, **13**, 130–43.

WARREN, R. M. (1958). A basis for judgments of sensory intensity. *Amer. J. Psychol.*, **71**, 675–87.

WARREN, R. M., and POULTON, E. C. (1960). Basis for lightness-judgments of grays. *Amer. J. Psychol.*, **73**, 380–7.

WARREN, R. M., SERSEN, E. A., and PORES, E. B. (1958). A basis for loudness-judgments. *Amer. J. Psychol.*, **71**, 700–9.

WARREN, R. M., and WARREN, R. P. (1956). Effect of the relative volume of standard and comparison-object on half-heaviness judgments. *Amer. J. Psychol.*, **69**, 640–3.

WARREN, R. M., and WARREN, R. P. (1958). Basis for judgments of relative brightness. *J. opt. Soc. Amer.*, **48**, 445–50.

WEVER, E. G. (1949). *Theory of Hearing*. New York.

*Manuscript received* 11*th May*, 1962.

# THE EFFECT OF INSTRUCTIONS ON THE JUDGEMENT OF BRIGHTNESS

BY

A. A. LANDAUER

*From the Department of Psychology, University of Sydney*

A brightness constancy experiment using ten subjects showed that different instructions significantly affect the subject's PSE judgement. Variations in the reflectance of the standard target and in the background and illumination of the variable target did not influence the effect of different instructions. The use of the D-ratio as constancy index in this study confirms an earlier theoretical finding that other brightness constancy ratios introduce an artefact into experimental results.

## INTRODUCTION

The response of subjects in the performance of a perceptual task is to some extent determined by the verbal instructions given. The common view seems to be that instructions serve to induce "analytic" or "object-directed" viewing "attitudes" which considerably affect the perceptual behaviour of the observer. A study of the importance of the instruction stimulus in brightness constancy has been reported by Henneman (1935). Earlier investigations on the effect of instructions and "induced set" on the degree of other constancies were made, amongst others, by Thouless (1932) and Klimpfinger (1933).

Henneman (1935), who reported his results in terms of the E-ratio, an adaptation of the Thouless ratio, found that brightness constancy changed significantly in the majority of his subjects according to the type of instructions given. In a recent paper, Landauer (1962) argued that neither the Brunswik nor the Thouless ratio are applicable in comparative studies of brightness constancy, and that the D-ratio, or any other measure developed on the same principle, appears to be the only legitimate method to report and compare experimental findings. Similar views have been expressed by Leibowitz (1956) and McDonald (1962).

The D-ratio is the ratio of the product of incident light and reflectance for the standard to the same product for the variable at the PSE. Reflectances are expressed in terms of the Munsell value scale and are obtained from the table developed by Newhall, Nickerson and Judd (1943). These range from absolute white (Munsell *value* 10) to absolute black (Munsell *value* 0). The amount of incident light is expressed as a percentage of a fixed, arbitrary constant and it is not necessary to convert this into any of the standard illumination measuring units.

To compute the D-ratio the reflectance values of the standard and of the variable at the PSE are empirically determined. If adjustable revolving black and white discs are used as stimuli, only the reflectance of these discs needs to be empirically determined and the reflectance of the experimental settings can then be calculated. The reflectance value of each stimulus is multiplied by the amount of incident light to obtain the luminance value of each stimulus. The D-ratio is then determined by dividing the luminance of the variable by the luminance of the standard.

The D-ratio permits comparison between various experimental conditions; this cannot always be done when either the Brunswik or the Thouless ratio is used. For instance in a perfect constancy situation where the difference between the levels of illumination of the two stimuli approaches the differential threshold, the older ratios give the same numerical value as in a situation where the subject shows perfect constancy under widely different levels of illumination. In addition, an examination of the

Brunswik and Thouless ratios, using the data of Hsia (1943) shows that they are perfectly correlated, provided that the measured physical conditions from which the ratios have been computed have a constant relationship. Thus the two ratios will be perfectly correlated when the illumination of the standard and the illumination of the variable are in constant relationship. If, however, this relationship is changed, neither the Brunswik nor the Thouless ratios represent the same scale.

When the Brunswik or the Thouless ratio is zero there is no constancy and when these ratios are unity there is perfect constancy. In the D-ratio the greatest departure from unity occurs when there is a large difference between the luminance of the standard and of the variable. When there is little difference the ratio will approximate unity, and at perfect luminance equality (zero constancy) the ratio will be unity. Except in cases of "over-constancy" the older ratios can never exceed unity, while the D-ratio in instances of extreme differences in the levels of illumination and where "constancy" is shown, departs from unity and shows, moreover, a wider spread of scores. Such a departure from unity is in the D-ratio always a function of the ratio of the levels of illumination of the standard and the variable. In cases where the standard has the higher level of illumination, the value of the D-ratio at "complete constancy" is equal to the ratio of the levels of illumination pertaining on the standard and the variable. In cases where the standard has the higher level of illumination, the D-ratio at complete constancy is the reciprocal of the ratio of the levels of illumination of the two targets. Provisions can be made in the computation of the D-ratio for differences in the viewing distance of the stimuli and for equating different j.n.d. levels of luminance. However, as far as the application of the D-ratio is concerned, its main advantage lies in the fact that it permits the comparison of data of experiments in which the relationship of illumination conditions between the stimuli is varied.

In the absence of any reliable method of assessing subjective states such as "perceptual attitude" or "induced set," it is preferred here to observe perceptual behaviour as a function of the instructions given, without speculating on the nature of these states.

It is well known that in making brightness matches an observer seldom achieves physical luminance equality between standard and variable. This discrepancy between physical and subjective equality derives from a number of factors which include adaptation, contrast effects, the tendency towards brightness constancy as well as the effect of instructions on judgement. It is with this latter determinant that this study is mainly concerned.

A number of tentative hypotheses have been advanced suggesting various causes for the effect of the instruction stimulus. The most common determinant mentioned is the "familiarity" of the observer with the stimulus. The familiarity factor derives largely from the memory colour hypothesis of Hering (1874) and, it has been conclusively shown by numerous studies (cf. Duncker, 1939) that brightness constancy is increased in the PSE judgement if the subject "believes" that he is looking at a familiar object whose reflectance he "knows."

If, however, unfamiliar and relatively meaningless stimuli are presented to the subjects, and instructions only are varied, an explanation in terms of previous knowledge of the reflectance of the stimulus cannot be applied. Since Henneman's (1935) results are not very convincing and are, moreover, questionable in terms of the constancy index used, it was decided to perform an experiment in which not only instructions, but also such stimulus properties as reflectance of the standard target, differences in the levels of illumination between the standard and the variable target, and differences in the reflectance and hue of the background of the standard target, were systematically varied. These variables were investigated since it seemed reasonable to suppose that instructions would significantly modify the subject's response, and that this effect would be independent of any other stimulus properties which were varied. That is, the "attitudinal" state induced by instructions would operate irrespective of any variations in stimulus properties which were introduced. Such independence would be revealed by the absence of any significant interaction between instructions and other stimulus properties.

## APPARATUS

A rectangular hardboard box was divided into two equal lightproof compartments, each of which was 88·5 cm. long, 47·2 cm. high and 45·8 cm. wide. The interior of each compartment, except the lids and rear walls, were covered with glossy white paper. The rear wall of the compartment on the subject's left (the standard target side) was covered with white cardboard of 90·5 per cent. reflectance, and on the rear wall of the compartment on the subject's right (the variable target side) provisions were made to attach cardboard backgrounds of different reflectances and hues.

Through the rear wall of the box the shafts of two electrically-driven colour mixers protruded 5 cm. into each compartment. Cardboard discs of a diameter of 20·5 cm. could be mounted on both shafts which were slanted at an angle of 15° upwards from the true horizontal. The rotation and the angular presentation of the targets was introduced in order to give a textureless "aperture colour" effect. The variable colour mixer on the right permitted the relationship between two interlocking cardboard discs to be changed by the experimenter while the discs were rotating. The angle of the sectors of the interlocking discs could be read from a dial. Discs of different reflectances were used on the standard target and these were made from black and white cardboard having a reflectance of 90·5 and 2·5 per cent. respectively. The same cardboards were used to make the interlocking discs of the variable target.

A pair of viewing tubes was affixed to the front wall of the box; one leading into each compartment. These tubes were 10 cm. long, had a diameter of 25·5 mm., and each was set at 12·5° outwards. This arrangement gave a clear view of the targets and of a part of the rear walls, the standard through the left tube and the variable through the right. The interior of the compartments were not visible to the subject. An artificial pupil with a diameter of 2·5 mm. was placed in each tube 5 mm. from the eye. Eye pieces were mounted on the viewing tubes to prevent the entry of stray light.

The diameter of each viewing tube aperture could be reduced to 11 mm. (restricted viewing condition) or closed by means of a pair of sliders which were affixed to the inside front wall of the box. The visual angle of the targets and their surround was 14·2° under normal viewing (diameter of the viewing tubes), and 5·1° when part of the target only was seen under restricted viewing. The distance from the centre of the targets to the eyes was 93·7 cm.

A frosted 200-watt incandescent globe provided the illumination for each compartment. The globes were situated in a housing on top of the box and an iris diaphragm with a maximum aperture of 24 mm. controlled the amount of light entering each compartment. The maximum aperture of the diaphragm was arbitrarily fixed as unity for illumination. The apparatus was placed on a table with the viewing tubes at eye level for a seated subject.

## SUBJECTS

A total of 10 male and female staff members or senior students of the Department of Psychology served as subjects. Three subjects had corrected vision but did not wear glasses during the observations.

## INSTRUCTIONS

Two sets of instructions were recorded on tape. Since these represent the principal variable in this study they are given in full.

*Instructions* 1: "It is extremely important that you follow the instructions exactly. Presently you will see through those viewing tubes in front of you two patches of grey colour. I can adjust the right patch to exactly the same shade as you tell me. As you know, grey is a mixture of black and white, and as I keep on adjusting the shade of grey you just tell me either 'blacker' or 'whiter,' as the case may be, and tell me when the two shades are exactly alike. I want you to consider only the greyness, the amount of black and white, in the two blotches of grey and disregard the illumination. Remember, a white object is white even if it is in the shade, and a black object is black, even if it is in bright light. You will find it easier when making your judgement if you go first past the point of equality and then tell me to go back, and by going forward and backwards a few times reach the point in which you get the best match. I shall show you exactly what I mean. Please look now through the tubes. The two patches of grey are exactly alike. Now I am increasing the illumination of the right patch—and now I am decreasing the illumination—but I have not changed the shade of the disc. It remains the same.

After you have made your judgement the tubes will be blacked out till a new shade matching is to be made. Try not to move around during the actual matching and make your judgement as fast as you can."

*Instructions* 2: "It is extremely important that you follow the instructions exactly. Presently you will see through those viewing tubes in front of you two patches of grey. I can adjust the right patch to exactly the same shade as you tell me. You know that the brightness of an object depends upon the amount of light and its colour, and that the greys which you will see are a mixture of light and dark just as much as they are of black and white. For instance on a photograph a light grey in the shade looks darker than a dark grey which is in the light. As I keep on adjusting the shade of grey I want you to tell me either 'darker' or 'lighter,' as the case may be, and tell me when the two shades look exactly alike. You will find it easier when making your judgement if you go first past the point of equality and then tell me to go back, and by going forward and backwards a few times reach the point in which you get the best match. I shall show you exactly what I mean. Please look now through the tubes. The two patches of grey are exactly alike. Now I am decreasing the amount of light on the right side. Obviously the right side looks darker now. So I am now putting more white into the mixture. This approximately will make the right side again equal to the other side; they have the same brightness as before. After you have made your judgement the tubes will be darkened till a new shade matching is to be made. Try not to move around during the actual matching and make your judgement as fast as you can."

Before each trial the subjects were given either of the following reminders: *Instructions* 1: "Don't forget, you have to match the shades of grey only and disregard the illumination." *Instructions* 2: "Don't forget you have to match the patches of grey for their brightness, accounting both for their greyness and illumination."

## EXPERIMENTAL PROCEDURE

The standard target side was always presented with full illumination and under restricted (reduced) viewing condition, i.e. it appeared against a black surround. The following conditions were systematically varied:

A. *Instructions*: Either instructions 1 or instructions 2 were given to the subject at the beginning of each session. While the instructions were played to the subject the demonstrations implied in them were given.

B. *Reflectance of Standard Target*: In any one session the standard target had either a reflectance of 20, 10 or 5 per cent.

C. *Illumination of the Variable Target Compartment*: In each session both levels of illumination of the variable target compartment were used. This was either 50 or 25 per cent. of the level of illumination obtaining in the standard target compartment.

D. *Surround of the Variable Target*: In each session all four of the following surrounds were used:

(i) white, reflectance 90·5 per cent.;

(ii) red, reflectance 16 per cent. at 670 m$\mu$;

(iii) blue, reflectance 16 per cent. at 460 m$\mu$;

(iv) restricted viewing, i.e. with only part of the variable seen against a black surround.

Under all conditions the dependent variable was the setting which the subject directed the experimenter to make on the variable colour mixer.

Each subject attended six sessions held on different days. The order of the sessions was randomized over the two different sets of instructions and over the three levels of reflectance of the standard target. Each session consisted of eight trials in which the subject was presented in random order with each of the four viewing conditions at each of the two levels of illumination. Each trial consisted of three matching judgements which were made consecutively both in ascending (from black to white) and descending order. The observations took place in a room lit by fluorescent light: each session lasted about twenty minutes.

The dial reading of each match was converted into reflectance values and averaged over each trial. The mean reflectance value of each trial was multiplied with the amount of light obtaining and this value constituted the luminance of the target. The luminance values of both the standard and the variable targets were used to calculate the D-ratio (Landauer, 1962) which served as the constancy index.

## RESULTS AND DISCUSSION

TABLE I

| *Source* | *S.S.* | *d.f.* | *M.S.* | *F* | *Probability* |
|---|---|---|---|---|---|
| Instructions .. .. .. .. | 8·8726 | 1 | 8·8726 | 11·08 | $0·005 < p < 0·01$ |
| Error for Instructions .. .. | 7·2049 | 9 | 0·8005 | | |
| Reflectance of Standard Target .. | 23·0948 | 2 | 11·5474 | 7·66 | $0·001 < p < 0·005$ |
| Error for Reflectance .. .. | 27·1486 | 18 | 1·5083 | | |
| Illumination of the Variable Target Compartment .. .. .. | 10·6803 | 1 | 10·6803 | 28·48 | $p < 0·001$ |
| Error for Illumination .. .. | 3·3752 | 9 | 0·3750 | | |
| Surround of the Variable Target .. | 27·7453 | 3 | 9·2484 | 18·63 | $p < 0·001$ |
| Error for Surround .. .. .. | 13·4051 | 27 | 0·4965 | | |
| *Interaction:* Reflectance × Surround | 13·3045 | 6 | 2·2174 | 10·69 | $p < 0·001$ |
| Error for Interaction: Reflectance × Surround .. .. .. .. | 11·2056 | 54 | 0·2075 | | |
| *Interaction:* Illumination × Surround | 7·7112 | 3 | 2·5704 | 22·12 | $p < 0·001$ |
| Error for Interaction: Illumination × Surround .. .. .. .. | 3·1387 | 27 | 0·1162 | | |
| *Residual* .. .. .. .. .. | 122·0395 | 319 | | | |
| Total .. .. .. .. | 278·9263 | 479 | | | |

An analysis of variance of the data is summarized in Table I. To obtain the greatest amount of information from the data it was decided not to pool interactions, but for reasons of parsimony only the significant effects and their respective error terms are reported here. A probability level of less than 5 per cent. ($p < 0·05$) was arbitrarily fixed as a significant difference level. The error term in all the F-ratios was the variance estimate of the next highest interaction effect involving subjects.

A. *Instructions:* For all subjects and under all conditions the mean D-ratio of the PSE judgement was 1·236 with instructions 1, and 1·507 with instructions 2. This difference is significant and the absence of any significant higher-order interactions involving the instruction stimulus confirms the hypothesis that instructions significantly affect a subject's PSE judgement in a brightness match. The results also confirm that this effect is, within the limits of the experimental conditions, independent of the reflectance and background of the variable target as well as of its level of illumination.

B. *Reflectance of Standard Target:* For all subjects and under all conditions the mean D-ratio of the PSE judgement was 1·092 when the reflectance of the standard target was 20 per cent., 1·395 when the reflectance of the standard target was 10 per cent., and 1·628 when the reflectance of the standard target was 5 per cent. The difference between the mean response given to targets with a reflectance of 20 per cent. and either 10 or 5 per cent. was significant. (An adaptation of Lindquist's (1953) formula 20 was used to determine the significance of differences when more than two means were involved). The significant interaction between the reflectance of the standard target and the background of the variable target is due to the fact that the D-ratio tends to reach unity when the surround of the standard and variable target are equal. This confirms Wallach's (1948) relational theory.

C. *Illumination of the Variable Target Compartment:* For all subjects and under all conditions the mean D-ratio of the PSE judgement was 1·521 when the level of illumination of the variable target compartment was 50 per cent., and 1·222 when the level of illumination of the variable target compartment was 25 per cent. This highly significant difference is contrary to Hsia's (1943) finding that brightness constancy is increased as the absolute difference in illumination between the two targets is increased. It confirms Leibowitz's (1956) assumption that Hsia's finding is due to an artefact of the ratio which he used. The significant interaction between the level of illumination and surround of the variable target is due to the fact that when the surround of both the standard and the variable target become equal, a judgement close to physical equality is given. That is, that little constancy is manifested.

D. *Surround of the Variable Target:* For all subjects and under all conditions the mean D-ratio of the PSE judgement was 1·770 when the variable was matched against a white surround, 1·347 against a red surround, 1·214 against a blue surround of equal reflectance to that of the red, and 1·155 against a black surround. The mean match made against a white surround is significantly different to the match made against other backgrounds. A significant difference occurs between the mean matches made against a red surround and similar judgements with a black surround. No significant difference, however, occurs between blue and black surrounds. The mean D-ratio made with the red surround tended to be greater than that with a blue surround of equal reflectance. This difference, however, was not significant. It appears therefore that the two colours should have been equated for phenomenal rather than for physical reflectance. A similar finding has been reported by Graham (1959) who found that red and yellow hues seem to have a brightening effect, while greens and blues seem to have a darkening effect.

## CONCLUSION

This study has clearly confirmed that, under the experimental conditions detailed, brightness matches are determined, *inter alia*, by (1) the instructions given to the subject; (2) the relation of a target's reflectance to the reflectance of its surround; and (3) the difference in the levels of illumination of the two surfaces which are compared. The use of the D-ratio has confirmed Leibowitz's (1956) theoretical argument that the paradoxical findings of Hsia (1943) are due to an artefact of the ratio employed.

The assistance of Dr. R. H. Day in the planning of this study, and of Mr. J. E. Shaw of the National Standard Laboratory of Australia, in making photometric measurements, is gratefully acknowledged.

## REFERENCES

DUNCKER, K. (1939). The influence of past experience upon perceptual properties. *Amer. J. Psychol.*, **52**, 255–65.

GRAHAM, C. H. (1959). Color theory. In KOCH, S. ed., *Psychology: A Study of a Science* (Vol. 1). New York.

HENNEMAN, R. H. (1935). A photometric study of the perception of object color. *Arch. Psychol. N.Y.*, **27**, No. 179.

HERING, E. (1874). Zur Lehre vom Lichtsinne (Vierte Mittheilung). *S.B. Akad. Wiss. Wien. Abt. III*, **69**, 85–104.

HSIA, Y. (1943). Whiteness constancy as a function of difference in illumination. *Arch. Psychol. N.Y.*, **40**, No. 284.

KLIMPFINGER, S. (1933). Über den Einfluss von intentionaler Einstellung und Übung auf die Gestaltkonstanz. *Arch. ges. Psychol.*, **88**, 551–628.

LANDAUER, A. A. (1962). The D ratio: a new method for the measurement of brightness constancy. *Aust. J. Psychol.*, **14**, 22–6.

LEIBOWITZ, H. (1956). Relation between the Brunswik and Thouless ratios and functional relations in experimental investigations of perceived shape, size and brightness. *Percept. mot. Skills*, **6**, 65–8.

LINDQUIST, E. F. (1953). *Design and Analysis of Experiments in Psychology and Education*. Boston.

McDONALD, R. P. (1962). An artifact of the Brunswik ratio. *Amer. J. Psychol.*, **75**, 152–4.

NEWHALL, S. M., NICKERSON, D., and JUDD, D. B. (1943). Final report of the O.S.A. Subcommittee on the spacing of the Munsell colors. *J. opt. Soc. Amer.*, **33**, 385–418.

THOULESS, R. H. (1932). Individual differences in phenomenal regression. *Brit. J. Psychol.*, **22**, 216–42.

WALLACH, H. (1948). Brightness constancy and the nature of achromatic colors. *J. exp. Psychol.*, **38**, 310–24.

*Manuscript received 2nd June, 1962.*

# THE EFFECT OF SELF-CONTRADICTION ON FALLACIOUS REASONING

BY

P. C. WASON

*From the Medical Research Council Industrial Psychology Research Unit,*
*University College, London*

A serial task was used in which the possibility of making related valid and fallacious inferences was alternated over a series of trials. The hypothesis was investigated that if valid inferences are forced to contradict previously made fallacious inferences, then subsequent fallacious inferences will be withheld. The results showed that self-contradiction tended to extinguish fallacious inferences in an all-or-none manner without affecting valid inferences.

## INTRODUCTION

In deductive reasoning a valid inference follows necessarily from its premises, i.e. it would be inconsistent to assert the premises and deny the inference. A fallacious inference is an invalid inference which has the appearance of validity. An individual who commits a fallacy is being deceived into accepting a conclusion which merely looks correct. For example, an individual presented with the statement, "all A are B" might suppose that "all B are A" follows from it. On the other hand, this mistake would not be made if a statement of the same form had been used to express a known fact, e.g. "all monkeys are animals." The abstract nature of the material about which inferences are made is one variable which affects the tendency to reason fallaciously (Wilkins, 1928); another is the emotive quality of the material (Lefford, 1946). The aim of the present investigation, is to make the individual aware of his deception by forcing valid inferences to be inconsistent with previously made fallacious inferences in a serial task. This obliges the individual to contradict himself. The hypothesis is investigated that this self-contradiction reduces the tendency to make further fallacious inferences of the same type.

## METHOD

### *The fallacies and the tasks*

An attempt was made to induce two logically equivalent fallacies connected with conditional statements. In order to explain them the logic of conditionals will be briefly discussed. A statement in the form "if p then q" is false only in the case in which the antecedent p is true and the consequent q is false. If this conditional statement is true, it follows that (i) if p is true, then q must be true, and (ii) if q is false, then p must be false. However, to deny the antecedent, i.e. to infer that if p is false, then q must be false, or to affirm the consequent, i.e. to infer that if q is true, then p must be true, is to be guilty of a fallacious inference.

Two tasks were used: an "antecedent task," designed to induce the denial of the antecedent ("if not-p then not-q"), and a "consequent task," designed to induce the affirmation of the consequent ("if q then p"). The following incomplete rules were used in the antecedent and consequent tasks respectively:

"Any employee aged 34 years, or more, will receive a salary of at least £—— a year."
"Any employee aged —— years, or more, will receive a salary of at least £1,900 a year."

In the antecedent task it would be valid to infer that the "critical salary," i.e. the salary omitted in the rule, *could not be* more than that received by any employee aged at least 34, but fallacious to infer that it *must be* more than that received by any employee under 34. In the consequent task it would be valid to infer that the "critical age," i.e. the age omitted in the rule, *must be* more than that of any employee receiving less than £1,900, but fallacious to infer that it *could not be* more than that of any employee receiving at least £1,900.

The incomplete rule for the antecedent, or the consequent task, was presented, followed by the ages and salaries of ten hypothetical employees, making a series of ten trials. These particulars allowed valid inferences on the odd numbered trials and fallacious inferences on the even numbered trials. At each trial the subjects were required either to make, or withhold, a deductive inference about the possible limits of the critical salary (age), and then state their current estimate of it. The estimates were intended to motivate the subjects and provide a check on the inferences.

*Design*

Sixty subjects (psychology undergraduates) were allocated at random before the experiment started to four groups: an experimental and a control group for each task. Inconsistency between fallacious and valid inferences was introduced in the experimental groups, but not in the control groups. In the antecedent task control group the salaries received by employees over 34 are all higher than the salaries received by employees under 34. It follows that successive valid and fallacious inferences about the limits of the critical salary are not forced to contradict each other. In the experimental group, however, the first four trials are the same as those in the control group, but thereafter the salaries given decrease successively. For example, on trial 4 an employee aged 24 receives £1,800, but on trial 5 an employee aged 37 receives £1,600. Hence the valid inference on trial 5, that the critical salary could not be more than £1,600, is inconsistent with the fallacious inference on trial 4 that the critical salary must be more than £1,800. But on trial 6 an employee aged 25 receives £1,500. The fallacious inference on this trial is consistent with the valid inference on trial 5. Thus the valid inferences on trials 5, 7 and 9 are forced to be inconsistent with the fallacious inferences which *precede* them, but are consistent with the fallacious inferences which *succeed* them. In the consequent task the particulars given in the control group are similarly arranged so that valid and fallacious inferences are consistent. In the experimental group inconsistency is again introduced on trial 5. For example, on trial 4 an employee receiving £2,300 is aged 35, but on trial 5 an employee receiving £1,500 is aged 37. The valid inference on trial 5, that the critical age must be more than 37, contradicts the fallacious inference on trial 4 that the critical age could not be more than 35. On the remaining trials the pattern of consistent and inconsistent inferences follows that in the antecedent task.

At every trial one type of question was asked in order to discriminate a valid inference, and another type to discriminate a fallacious inference. Hence one question was redundant at each trial. In the antecedent task the following types discriminated valid and fallacious inferences respectively: "*Could* x *be less than the critical salary?*" and "*Could* x *be more than the critical salary?*" (where x = the salary received by an employee on any given trial). In the consequent task the corresponding types of question were: "*Could* x *be over the critical age?*" and "*Could* x *be under the critical age?*" (where x = the age of an employee on any given trial). A negative answer to these modal questions indicates that an inference has been made. Such an answer rules out one of two possibilities, either (i) that a given salary (age) is less than the critical salary (age), or (ii) that it is more than the critical salary (age). An affirmative answer, on the other hand, is consistent with the truth of either possibility. Hence it is the correct answer when the information allows a fallacious inference: it constitutes a refusal to make one.

*Procedure*

Subjects were tested individually and instructed as follows: This is a reasoning task which I want you to think about carefully.

In a certain organization the following rule held without exception: (here the rule for the appropriate task was given). *Your task is to try to find out the critical salary* (*age*) *in the rule*. I shall show you the ages and salaries of a number of employees in the organization, one at a time. After examining the particulars of each, you will answer two questions about the critical salary (age), and then write down your estimate of it. You will not be timed. Have you any questions?

The rule was displayed on a card in front of the subject, and then the particulars of the first trial, and the two questions, were presented on a slip of paper. If the subject failed to answer the relevant question correctly, i.e. failed to make a valid inference, his response was queried by the experimenter who waited until the correct answer was given. This procedure was carried out only on the first trial. After the subject had recorded his estimate, the particulars of the second trial were presented, those of the first remaining on the table in front of the subject. When the second trial had been completed, the particulars of the first were removed, and those of the third presented. This procedure continued throughout the remaining trials. Thus after the first trial the subjects were able to consult the particulars, the answers and the estimates of each immediately preceding trial. The subjects were allowed to make notes on rough paper provided, but were not allowed, of course, to correct their responses to any previous trial.

## RESULTS

### *Consistency between answers and estimates*

The estimates of the critical salary and the critical age were consistent with the inferences (whether valid or fallacious) on 412 out of a possible 419 trials. This suggests that, regardless of whether the subjects were reasoning fallaciously, they were not random in their behaviour. On the remaining 181 trials inferences were withheld.

### *Initial susceptibility to fallacious inference*

The criterion for initial susceptibility to making the fallacies was the occurrence of fallacious inferences on trials 2 and 4. (Twenty-seven out of the thirty subjects, classified as not susceptible, withheld fallacious inferences on both these trials.) In the antecedent task ten out of thirty subjects were susceptible, and in the consequent task twenty out of thirty subjects were susceptible. These frequencies differ significantly from the expected ones ($\chi^2 = 5{\cdot}4$, $p < 0{\cdot}025$, two tail test) and indicate that the affirmation of the consequent is a much more deceptive fallacy than the denial of the antecedent.

### *Effects of self-contradiction*

Table I shows the number of subjects, initially susceptible to making fallacious inferences, who withhold subsequent fallacious inferences at different stages of the tasks, or make the maximum number of fallacious inferences, i.e. make no correction, in the experimental and control groups in each task. Once one of these subjects had withheld such an inference, they never succumbed to making one again. The differences between the distributions of the experimental and control groups within each task were tested by Whitfield's dichotomous tau (Whitfield, 1947). In the antecedent task the chance probability of the difference between the groups = 0·05; in the consequent task the chance probability = 0·08 (one tail tests). In spite of the significant difference between the tasks (ten subjects susceptible to making fallacious inferences in the antecedent task and twenty in the consequent task), it is clear that the effects of self-contradiction are similar in both. When the S values and variances of the tasks are combined, the chance probability of a difference = 0·03.

Four subjects failed to make valid inferences on subsequent trials, one failing on three trials, and three failing on a single trial. Only one of these subjects was in an experimental group. Thus it is evident that, although self-contradiction occurs in making a valid inference, it did not cause *these* inferences to be withheld. Nine of the total subjects had to be corrected on the first trial, four in the antecedent task and five in the consequent task.

TABLE 1

NUMBER OF SUBJECTS, INITIALLY SUSCEPTIBLE TO MAKING FALLACIOUS INFERENCES, WHO RESIST MAKING SUBSEQUENT FALLACIOUS INFERENCES AT DIFFERENT STAGES OF THE TASKS, OR MAKE THE MAXIMUM NUMBER OF FALLACIOUS INFERENCES, I.E. MAKE NO CORRECTION

| | *Antecedent task* | | *Consequent task* | | *Combined task* | |
|---|---|---|---|---|---|---|
| | *Experimental Group* | *Control Group* | *Experimental Group* | *Control Group* | *Experimental Group* | *Control Group* |
| *N* | 4 | 6 | 10 | 10 | 14 | 16 |
| 6th Trial .. | 3 | 0 | 5 | 2 | 8 | 2 |
| 8th Trial .. | 0 | 1 | 1 | 0 | 1 | 1 |
| 10th Trial .. | 0 | 1 | 0 | 0 | 0 | 1 |
| No correction .. | 1 | 4 | 4 | 8 | 5 | 12 |

## DISCUSSION

In this exploratory experiment the attempt was made to induce fallacious inferences initially by using tasks which embodied a familiar situation: a positive correlation between age and salary. It is a plausible inference (other things being equal) that older employees *probably* earn more than younger ones. This inference, however, is inductive but the questions posed in the tasks could elicit only deductive inferences, i.e. inferences which are certain rather than probable. But it seems likely that the habitual tendency to reason inductively was responsible for the fallacious inferences to some extent. A similar interpretation of errors made in syllogistic reasoning has been proposed by Chapman and Chapman (1959).

In the experimental groups self-contradiction occurs when a valid inference is forced to be inconsistent with previously made fallacious inferences. The results show that these valid inferences are not withheld, but that subsequent fallacious inferences (which are consistent with preceding valid inferences) do tend to be withheld. This suggests that inconsistency allowed the subjects to gain insight into the fallaciousness of their reasoning, rather than weakened the response associated with making a fallacious inference. In addition, it is clear that the effects of inconsistency tend either to work when it is first introduced, or not at all. In the combined experimental groups eight of the fourteen subjects (57·2 per cent.), initially susceptible to making fallacious inferences, withheld them at trial 6 (the first trial on which the effects of inconsistency could be felt), compared with two out of sixteen subjects (12·5 per cent.) in the combined control groups. But on subsequent trials there is little difference between the groups. In the experimental groups one of the six remaining subjects (16·7 per cent.) withheld fallacious inferences, compared with two out of fourteen (14·3 per cent.) in the control groups.

It has been shown that self-contradiction can be used to extinguish fallacious inferences, and it presumably does so by creating a *reductio ad absurdum* to remind the individual that he has inferred a conclusion which does not necessarily follow. But it is worth noting that a fallacious inference is a special kind of error. It is not analogous to a conceptual mistake in mathematics, for it is not necessarily the result of using a rule wrongly. Rules of reasoning are not taught as part of general education, nor are deductions about past events coded into an explicit system for future reference. Hence, whether the present method of correcting error has any generality, and whether it has any advantages over didactic correction (e.g. in "teaching machines") is a matter for future research.

Grateful acknowledgements are due to Dr. N. F. Dixon and Dr. Sheila Jones, both of whom read the manuscript, and offered invaluable critical comments, and to the Medical Research Council for a grant for scientific assistance.

## REFERENCES

CHAPMAN, L. J., and CHAPMAN, J. P. (1959). Atmosphere effect re-examined. *J. exp. Psychol.*, **58**, 220–6.

LEFFORD, A. (1946). The influence of emotional subject matter on logical reasoning. *J. gen. Psychol.*, **34**, 127–51.

WHITFIELD, J. W. (1947). Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, **34**, 292.

WILKINS, M. C. (1928). The effect of changed material on ability to do formal syllogistic reasoning. *Arch. Psychol.*, **102**.

*Manuscript received 26th May, 1962.*

# ON THE RESOLUTION OF SMALL TIME INTERVALS AND THE EFFECT OF CONDUCTION DELAYS ON THE JUDGEMENT OF SIMULTANEITY*

BY

A. M. HALLIDAY and ROSEMARY MINGAY†

*From the Medical Research Council, Institute of Neurology, National Hospital, Queen Square, London, W.C.1*

In the course of experiments on the cerebral evoked potential, recorded by a summation technique through the scalp, it was possible to measure directly the latency of the potential due to afferent conduction time. In two subjects the response following electrical stimulation of the toes was found to occur approximately 20 millisec. later than that following stimulation of the fingers. Using similar stimuli to finger and toe, parallel experiments have been carried out to determine whether there is any corresponding error in the judgement of when two stimuli are simultaneous.

The same subjects were presented with pairs of stimuli at a series of intervals up to 100 millisec. and made judgements as to whether they occurred at the same time, or if not, which was first. A control situation, in which the stimuli were given to the two index fingers, was compared with the experimental situation, in which stimuli were presented to finger and toe on opposite sides. The results showed a significant difference in the mean interval at which the two stimuli were judged simultaneous when the conduction pathways were of unequal length, and this difference was in the direction expected should the judgement of simultaneity depend on the simultaneous arrival of the afferent volleys centrally.

However, in only one subject was this difference of the order of 20 millisec., as might be expected were the relevant event the onset of the cortical evoked potential. Further, in the other subject the point of apparent simultaneity in the control situation, in which stimuli were given to the right and left index fingers, does not correspond with the point of true simultaneity, as would be expected with conduction pathways of equal length.

It is concluded, therefore, that while the temporal relationship of two stimuli is judged on the time of their central arrival rather than on the time of their occurrence at the periphery, judgements may be affected by factors other than conduction delays.

## INTRODUCTION

Using the summation technique devised by Dawson (1954) it is possible to record the response evoked in the cerebral cortex of a conscious human subject by an electrical stimulus given to hand or foot. The potential is recorded by an electrode placed on the scalp of the subject over that part of the primary sensory receiving area corresponding to the site of stimulation, with respect to an indifferent electrode placed more anteriorly on the head. It has a characteristic form, and begins after a latency representing the time taken for the sensory volley to travel from the point of stimulation to the cortex.

Examination of this latency in two subjects showed that for a similar electrical stimulus given to the hand and the foot, the onset of the first surface positive component of the cortical evoked potential occurred approximately 20 millisec. later following stimulation of the foot than following the hand stimulus. Thus the incoming sensory volley arrives at the cortex appreciably later when it has to traverse the longer afferent pathway.

---

* Based on a paper given to the E.P.S. in January, 1961.

† Present address: The Psychological Laboratory, Cambridge.

It appeared that further study of this situation might throw some light on the nature of the central perceptual mechanisms. If the perception of the time of occurrence of a stimulus depends on the arrival of the sensory volley at some particular part of the central nervous system in or near the cortex, and the volley from the foot takes approximately 20 millisec. longer, there appear to be two possibilities, viz. (*a*) that the stimuli are judged as together when the volleys arrive together centrally, and that in truly simultaneous stimuli, the foot stimulus will be judged slightly later in time than the hand stimulus; or (*b*) that the central mechanisms make allowance for the longer conduction times in the case of impulses from further away, and judge as simultaneous stimuli which actually arrive centrally separated by the appropriate interval. The latter alternative might be expected, for instance, if simultaneity were a learned concept, a particular temporal pattern of incoming impulses being associated with stimulus simultaneity.

The method adopted to investigate this problem was to give pairs of similar stimuli at intervals ranging from zero (i.e. simultaneity) to 100 millisec. by 5 millisec. steps, the subject in each case having to report whether the stimuli were together in time, or, if not, which was first. The two stimuli were given to the finger and toe on opposite sides of the body and a control situation was introduced in which the stimuli were given to the two index fingers.

It was predicted that, if the two stimuli were judged as simultaneous only when the volleys arrived at the cortex together, apparent simultaneity would not coincide with true simultaneity; the stimuli would appear together when the stimulus to the toe preceded that to the finger by approximately 20 millisec. However, because of the variability of a subject's resolution of time interval, there would be a scatter of "together" judgements about the point of apparent simultaneity. Any discrepancy due to conduction time should then appear as a shift in the *distribution* of "together" judgements, and the hypothesis would be supported if these judgements were distributed, not about true stimulus simultaneity, but about the stimulus interval corresponding to simultaneous cortical potentials.

## METHOD

### *Experimental design*

The *control* situation (in which the two stimuli were given to the index fingers) and the *experimental* situation (stimulation of index finger and contralateral second toe) were presented on alternate days to two adult subjects for ten days in all. Both subjects were right handed. For the experimental situation, the left index finger and right second toe were used for the first subject; the right index finger and left second toe for the second subject.

Four runs were given in each daily session, each run containing forty-one pairs of stimuli: one pair in each order (right or left stimulus first) at all intervals from 5 millisec. to 100 millisec. by 5 millisec. steps, and one simultaneous pair.

For the first two runs each day, the pairs of stimuli were presented in progressively decreasing intervals (with the stimulus on one side always leading), from 100 millisec. to simultaneity, and then in progressively increasing intervals with the other stimulus now leading. This situation in which the intervals were presented *sequentially* therefore corresponds to a determination of simultaneity by a method of decreasing and increasing limits. One of the two daily sequential runs began with the left stimulus first in the pair, the other run with the right first. To prevent the subject using the number of intervals presented as a clue to the position of simultaneity, a few extra 5-millisec. steps were introduced at the beginning and end of the sequential runs. They were not included in the calculation of results. This procedure and the equal number of runs given in either direction were designed to obviate the effect of the well-known "errors of anticipation."

For the third and fourth runs in each daily session, the same forty-one intervals were presented in random order. It was anticipated that the subjects would be less accurate in their judgements with this method of presentation.

The subject was told before each run whether the intervals were to be given sequentially or randomly. He was instructed to respond "Right first," "Left first" or "Together," but was allowed to give doubtful judgements in any of these categories.

### *Method of stimulation*

Stimulation was by means of condenser discharges (time constant 5–20 μsec.), provided by a two-channel thyratron stimulator, and was similar to that used in recording the cerebral evoked potentials. The stimuli were delivered through a silver disc electrode fixed with collodion to the pad of the index finger or second toe (terminal phalanx). A silver ring electrode, encircling the same digit more proximally, served as anode. Separate isolating transformers ensured the independence of the two channels.

The intervals between stimuli were timed by a digital delay generator of the type described by Pitman (1958). The particular unit employed enabled the stimuli to be given simultaneously or at any interval of 5 millisec. or a multiple thereof, each such interval being accurately timed to within 0·1 millisec. or better.

Sensory thresholds were determined at the beginning of each daily session, using the method of limits. The strengths of the stimuli were then doubled, and if the subject thought one stimulus was stronger than the other, one was adjusted until they felt subjectively equal. Subjects reported the sensation as somewhat like a touch; it was discrete in time, and did not outlast the instant of the stimulus pulse.

## RESULTS

### *Latencies of cortical evoked potentials*

The averaged cortical evoked potentials were recorded from both subjects. Measuring to the first inflexion of the initial surface positivity, the latencies were 23 and 42 millisec. for finger and toe respectively in Subject I, and 27 and 48 millisec. in Subject II. Sometimes an initial surface negativity can also be recorded, with a latency of a few millisec. less than the initial positivity. This negative wave has a very localized distribution on the cortex, and because of its variability, is not suitable for measurement. The true conduction time from periphery to cortex is therefore a few millisec. less than the latencies as measured to the beginning of the surface positive wave. It has been assumed that this will not lead to any appreciable error in calculating the *differences* between latencies. In the case of Subject I, the difference in latency between finger and toe was 19 millisec., for Subject II, 21 millisec.

### *Experiment on the judgement of simultaneity*

The results of the experiment in which the subject had to judge the temporal relationship of the stimuli give incidental information about the accuracy with which small time intervals can be resolved.

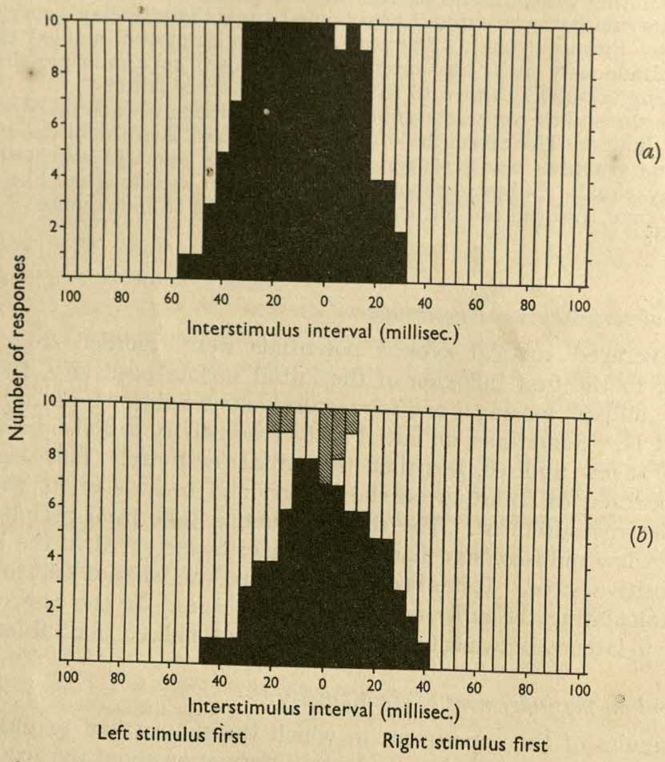
### *The control situation (stimulation of the two index fingers)*

Figures 1a and 2a show the distribution of correct order judgements, judgements of simultaneity, and incorrect order judgements for interstimulus intervals up to 100 millisec. on either side of simultaneity in the control situation. Doubtful judgements are included in the appropriate category. When the intervals were presented sequentially, Subject I made no errors about stimulus order, and Subject II made only a few errors, all of which occurred at intervals of less than 25 millisec. Pairs of stimuli were never judged as simultaneous by Subject I when over 55 millisec. apart, while Subject II seldom failed to resolve the order at intervals over 50 millisec.

When the pairs of stimuli were randomly presented, however, the resolution of interval and order was much less accurate; and both the number and distribution of judgements of simultaneity and incorrect order judgements were greater (Figs. 3a and 4a).

For Subject II the "together" judgements are distributed approximately equally on either side of true simultaneity, as was to be expected in the control situation (Figs. 2a and 4a). This is not the case, however, with Subject I, whose "together" judgements were made more often when the left stimulus preceded the right. This trend was significant in both sequential ($p < 0{\cdot}001$) and random ($p < 0{\cdot}01$) presentations (Figs. 1a and 3a). (Calculated from $t$ tests on the differences between the number of together judgements on either side of true simultaneity in each of the ten runs.) Cerebral evoked potentials recorded following stimulation of the two index fingers of this subject showed no measurable difference in latency. The discrepancy cannot, therefore, be attributed to a longer conduction time from the left index finger.

FIGURE 1



Distribution of responses given by Subject I to pairs of stimuli with the interstimulus interval shown on the abscissa. Pooled results of the ten runs in which the intervals were presented sequentially in 1 (*a*) for the control situation (stimuli to left and right index fingers) and in 1 (*b*) for the experimental situation (stimuli to left index finger and right second toe).

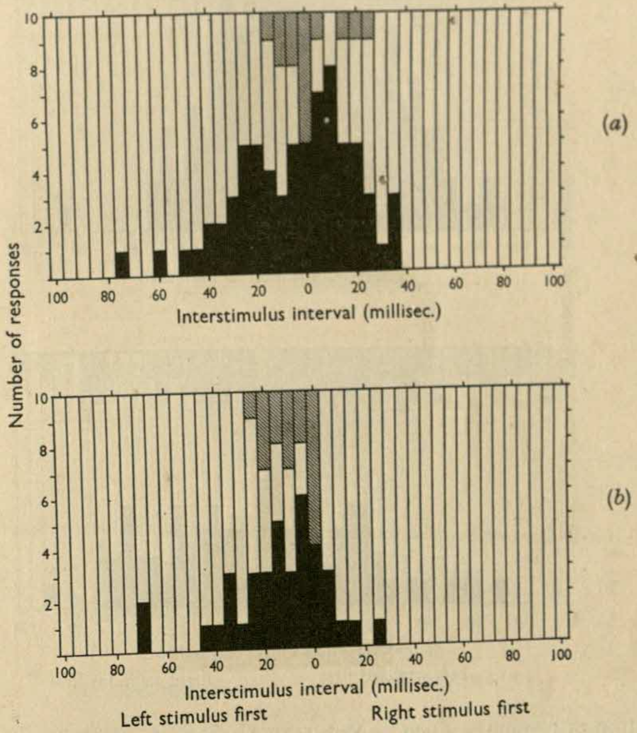
■ Stimuli judged together

▨ Order judged incorrectly

☐ Order judged correctly

In both subjects, the resolution of the order of the two stimuli was clearly much better in the sequential than in the random runs, and these differences are statistically significant (SI $p < 0{\cdot}001$, SII $p < 0{\cdot}01$, Figs. 1a and 2a compared with 3a and 4a). (Calculated from $t$ tests on the differences between the number of correct order responses in the sequential and random runs.) This is not surprising, since in the sequential runs the subject soon realizes the initial order of the stimuli and has only to determine two end-points, "togetherness" as simultaneity is approached, and "apartness" when the interval is increased again. In random runs, on the other hand, he has to make independent judgements as to the *order* of the two stimuli on each occasion when they are not judged to be simultaneous, without any clues from the preceding pairs. There were, therefore, more incorrect order judgements made in the random runs, but these were not always distributed symmetrically with respect to the "together" judgements. Subject II made more incorrect order judgements in the random runs when the right stimulus was first (Fig. 4a), but this trend was not

FIGURE 2



Distribution of responses given by Subject II to pairs of stimuli with the interstimulus interval shown on the abscissa. Pooled results of the ten runs in which the intervals were presented sequentially in 2 (*a*) for the control situation (stimuli to left and right index fingers) and in 2 (*b*) for the experimental situation (stimuli to left second toe and right index finger).

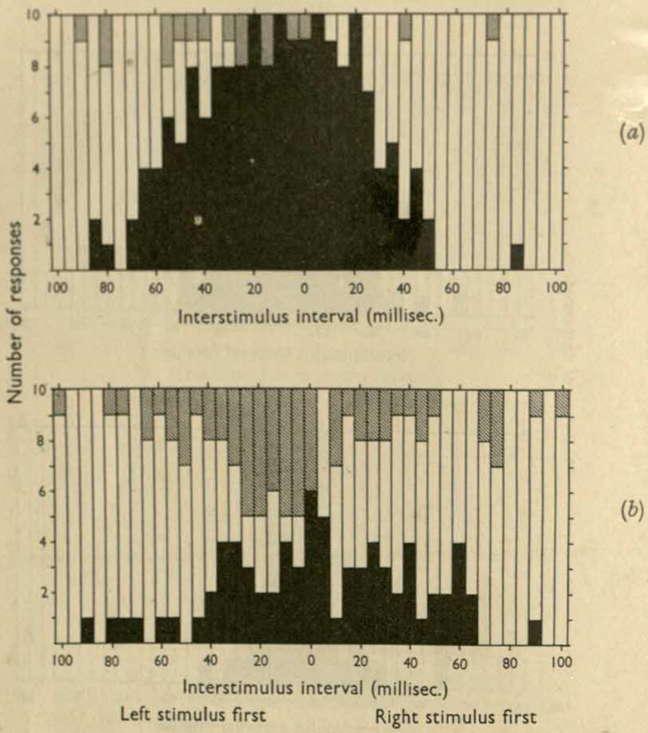
■ Stimuli judged together

▨ Order judged incorrectly

☐ Order judged correctly

significant ($p < 0{\cdot}1 > 0{\cdot}05$). (Calculated from $t$ tests on the differences betwee the number of incorrect order responses on either side of true simultaneity in each of the ten random runs.)

FIGURE 3



Distribution of responses given by Subject I to pairs of stimuli with the interstimulus interval shown on the abscissa. Pooled results of the ten runs in which the intervals were presented in random order in 3 (*a*) for the control situation (stimuli to left and right index fingers) and in 3 (*b*) for the experimental situation (stimuli to left index finger and right second toe).

■ Stimuli judged together
▨ Order judged incorrectly
☐ Order judged correctly

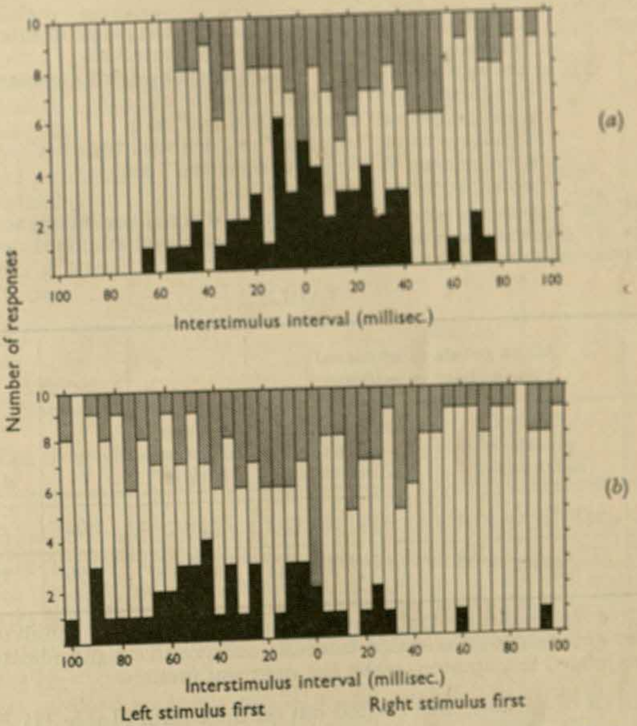
*The experimental situation* (*stimulation of finger and toe*)

The distribution of responses in the experimental situation is shown in Figures 1–4b, and can be compared with the corresponding control situation (Figs. 1–4a).

As in the control situation, both subjects were significantly better at resolving stimulus order in the sequential runs than in the random runs (SI $p < 0{\cdot}001$, SII $p < 0{\cdot}001$).

There is an increase in the number of incorrect order judgements when compared with the control situation, particularly in the random runs where these and the "together" judgements were distributed over the whole range of intervals (Figs. 3b and 4b) (SI $p < 0{\cdot}001$, SII $p < 0{\cdot}05$). (Calculated on the differences between the number of incorrect order responses in the random runs of the control and experimental situations.) In fact, at some of the shorter intervals, performance is virtually at chance level.

FIGURE 4



Distribution of responses given by Subject II to pairs of stimuli with the interstimulus interval shown on the abscissa. Pooled results of the ten runs in which the intervals were presented in random order in 4 (*a*) for the control situation (stimuli to left and right index fingers) and in 4 (*b*) for the experimental situation (stimuli to left second toe and right index finger).

■ Stimuli judged together ☐ Order judged correctly

▨ Order judged incorrectly

The number of "together" judgements is reduced for both subjects in both sequential (SI $p < 0{\cdot}01$, SII $p < 0{\cdot}05 > 0{\cdot}02$) and random runs (SI $p < 0{\cdot}001$, SII $p > 0{\cdot}1$). (Calculated on the differences between the number of "together" judgements in the control and experimental situations.) The smaller number of "together" responses makes the resolution of order appear better in the sequential runs than it did in the control situation. But the increase in incorrect order judgements in the random runs suggests that this appearance of better resolution in the experimental situation is an artefact, and that the results are related to a reduced tendency to see two stimuli as together when they are presented to finger and toe. Klemm (1925) noted a similar reluctance on the part of his subjects to ascribe simultaneity to stimuli given on widely separated parts of the body, e.g. on forehead and ankle.

### *The effects of conduction delays on judgements of simultaneity*

In order to test the original hypothesis of a shift in the distribution of judgements of simultaneity, it was necessary to calculate the point of apparent simultaneity in each run for each subject. To do this we used the following formula (Finney, 1952, Equation 20.10):—

$$m = x_k + \tfrac{1}{2}d - d\Sigma p_i$$

Where $x_k$ = 100 millisec., i.e. the upper limit of the range of intervals
$d$ = 5 millisec., i.e. the spacing between intervals
$\Sigma p_i$ = the sum obtained by scoring the subject's judgements as follows:—
a judgement of "left first" as 0
a judgement of "doubtful left first" as $\frac{1}{4}$
a judgement of "together" or "doubtful together" as $\frac{1}{2}$
a judgement of "doubtful right first" as $\frac{3}{4}$
a judgement of "right first" as 1.

In effect, "m" represents for *each* run, the interval in millisec. about which the subject's judgements of "together" are distributed. The means and standard deviations of these points of apparent simultaneity are given in Table I.

TABLE I

| | *Mean points of apparent simultaneity in millisec.* | | | | | |
|---|---|---|---|---|---|---|
| | *Control situation (index/index)* | *Experimental situation (index/toe)* | *Mean difference ("shift")* | *t* | *d.f.* | *p (one-tailed test)* |
| Subject I | −12·06 ± 4·04 | −2·94 ± 15·93 | 9·12 | 2·62 | 19 | <0·01 |
| Subject II | + 5·25 ± 14·68 | −12·19 ± 13·33 | 17·44 | 3·76 | 19 | <0·005 |

A negative sign indicates that the two stimuli were judged simultaneous when the left preceded the right by the stated interval; and positive sign indicates that the two stimuli were judged simultaneous when the right was first.

An analysis of variance was carried out on these data (Table II). It can be seen that the method of presentation of intervals (random or sequential) was not a significant cause of variance. Although these two psychophysical methods of presentation might have been expected to give different results, there was no significant difference between the mean "m" values for each subject for the random and sequential runs ($p > 0·1$ in all cases). This is borne out by the insignificance of the second order interaction "Subject × Presentation × Site." Apparently, therefore, the main effect of the different presentations is upon the variability of judgements and not upon the mean value. The results of the two methods of presentation have therefore been combined. The analysis of variance also indicates that there were no obvious practice effects between the first and second runs of each type of presentation. However, the interaction between the subjects and the experimental and control situations was significant at the 0·1 per cent. level.

For Subject II the mean difference of 17·4 millisec. ($p < 0·005$, Table I) represents a shift in the distribution of "together" judgements in the expected direction, i.e. the left stimulus (to the toe) must precede the right stimulus (to the index finger) for the subject to judge them as simultaneous. Subject I showed a shift of 9·1 millisec. in the opposite direction from Subject II, as predicted, and this difference between the control and experimental situations was also significant ($p < 0·01$).

In both subjects, therefore, significant shifts in the mean points of apparent simultaneity have been found, but in neither case were these of the magnitude predicted by the difference in latency of the cortical evoked potentials from finger and toe, viz. 19 millisec. for Subject I, 21 millisec. for Subject II. For Subject II apparent simultaneity for the control situation (+ 5·25 millisec.), with stimulation of the two index fingers and equal latencies to cortex, did not differ significantly from true simultaneity ($p > 0{\cdot}10$). But, for Subject I, apparent simultaneity for the control situation (− 12·06 millisec.) was found to be significantly different from true simultaneity ($p < 0{\cdot}001$).

TABLE II

Analysis of Variance

| *Source of variation* | *d.f.* | *Mean square* | *F ratio* | *p* |
|---|---|---|---|---|
| Subjects .. .. .. .. .. .. | 1 | 325·019 | 2·07 | >0·10 |
| Site of Stimulation (index/index or index/toe) | 1 | 345·488 | 2·20 | >0·10 |
| Presentation (Random or Sequential) .. | 1 | 7·051 | | |
| Runs (First or Second) .. .. .. | 1 | 72·676 | | |
| Subjects × Site of Stimulation .. .. | 1 | 3527·832 | 22·56 | <0·001 |
| Subjects × Presentation .. .. .. | 1 | 557·832 | 3·56 | >0·05 |
| Presentation × Site of Stimulation .. .. | 1 | 305·176 | 1·95 | >0·10 |
| Subjects × Runs .. .. .. .. | 1 | 43·144 | | |
| Site of Stimulation × Runs .. .. .. | 1 | 36·133 | | |
| Presentation × Runs .. .. .. .. | 1 | 3·301 | | |
| Site × Presentation × Runs .. .. | 1 | 325·020 | 2·07 | >0·10 |
| Subjects × Presentation × Site .. .. | 1 | 141·113 | | |
| Subjects × Site × Runs .. .. .. | 1 | 54·863 | | |
| Subjects × Presentation × Runs .. .. | 1 | 36·113 | | |
| Residual (including insignificant factors) .. | 73 | 156·372 | | |

A further test of the hypothesis can be made, taking into account these asymmetries in the control situation, and testing whether the observed points of apparent simultaneity for the experimental situation differ significantly from the values predicted by the differences in latency. Apparent simultaneity for the control situation is given by the "m" value for each run; we have assumed that the shifts in the experimental situation should have been 19 and 21 millisec. respectively in the appropriate direction, and have calculated the predicted points of apparent simultaneity for each of the corresponding experimental runs. Examining these differences by the $t$ test, it was found that the observed "m" values for Subject II in the experimental situation did not differ significantly from the value predicted ($p > 0{\cdot}4$), but for Subject I the "m" values were significantly different at the 1 per cent. level.

## Discussion

It can be concluded from these results that for both subjects differences in the length of conduction pathways alter the judgement of the temporal order of stimuli. These alterations are in the direction to be expected, if the perception of stimuli as simultaneous depends on their arrival together at some point in the central nervous system. No allowance therefore appears to be made for small differences in conduction time in judging the time of occurrence of stimuli.

On the other hand, in one subject the observed shift in judgements due to conduction delays did not correspond with the difference in the latency of the cortical evoked potentials. This might be due to the fact

(*a*) that there is some systematic error in measuring the latency of cortical evoked potential with respect to the relevant component,

(*b*) that the early cortical evoked potential in the specific area of the postcentral gyrus is not the relevant event on which temporal judgements are based, or

(*c*) that although judgements of temporal order depend primarily on the pattern of cortical firing initiated by the sensory volley, perception may vary with factors other than absolute time differences.

On the last-mentioned view, temporal judgements might be affected by "set" just as, for example, the way one sees a reversible figure such as Necker's cube is affected preferentially by a particular fixation point. The "prior entry" theory (Titchener, 1910), which suggests that when two stimuli occur at about the same time, the one which is being attended to will be seen as first, irrespective of the actual time relationship between the two, is one possible formulation of this type of hypothesis.

Certain features of the present results favour the third type of explanation. The interval at which the two stimuli were judged as simultaneous in the control situation did not correspond with true simultaneity in either subject, and in Subject I this difference was statistically significant. It seems possible that Subject I had a different "set" towards stimuli given to the right and left sides of the body. Another feature of the same kind was the incidence of incorrect order judgements made by Subject II, which were more numerous when stimuli were presented in a right-left order.

It is clear from the experiments of von Békésy (1957, 1959) that when stimuli are separated by only brief intervals, time differences at the periphery may be interpreted in terms of either temporal or spatial qualities. Two nearby stimuli to the fingertip may be felt as a single stimulus in different locations relative to the physical stimuli, or as two separate stimuli, and the sensation varies in a predictable way with small variations in the time relationships between the stimuli. Even widespread mechanical stimulation such as vibration of the whole forearm may under certain circumstances be interpreted as a quite sharply localized stimulus at one point in the stimulated area. This indicates that the idea of spatial localization of stimuli depending on point-to-point projection of body surface to cortex is too naïve, and that even the information contained in apparently elementary stimuli undergoes quite elaborate transformation and interpretation before reaching conscious awareness. The same is probably true of sensation of temporal differences. The present results show clearly that two volleys arriving at the same time cortically will not necessarily be regarded as simultaneous by the subject.

In the present experiments the spatial identity of the two stimuli would be expected to remain distinct because of their widely separated loci. However, the results in the random runs show that under these circumstances our subjects were extremely inaccurate in detecting the order in which they occurred, even when separated by intervals approaching a tenth of a second. These results can be compared with the findings of Ladefoged and Broadbent (1960), whose subjects were unable to judge accurately the time of occurrence of a click superimposed on a spoken sentence. The temporal acuity under the more favourable conditions of the sequential runs gave results comparable to those of earlier workers. Wundt (1903) gave a mean value of 27 millisec. as the longest interval at which successive stimuli fuse into a single experience. The sensation of more than one stimulus ("reduplication") was described by Golla and Antonovitch (1933) as appearing at intervals of 25–30 millisec., with stimulation of corresponding fingers of the right and left hands.

Hirsch and Sherrick (1961) have recently produced evidence that the interval needed to judge the order of two stimuli correctly in 75 per cent. of the trials is approximately 20 millisec., and that this is independent of the modality of the stimuli. In contrast, the interval required for judgements of "successiveness" (non-simultaneity) they found to vary with the modality employed, but to be considerably less than 20 millisec. For the tactile modality, they delivered taps to the tips of the two index fingers by means of an electro-mechanical transducer, and compared this with visual and auditory stimulation. They mention preliminary experiments in which they stimulated face and toe, finger and palm, or finger and arm, and required the subject to report on the order of stimuli, but they got variable and equivocal results, which they do not report further. They do not, therefore, give any information relative to simultaneity judgements with pathways of different length.

Stimulating different sites, Klemm (1925) found that a stimulus to the foot must precede one to the forehead by 25–30 millisec for the subject to judge them as simultaneous. He ascribed this interval to conduction time differences between foot and forehead. In the present experiments conduction delays clearly have an effect on the judgement of simultaneity, which confirms Klemm's conclusions. It is worth pointing out, however, that since he apparently made no comparable control observations on pathways of equal length, observations similar to those made on our Subject I in the experimental situation would have led him to the wrong conclusion.

It may be concluded that while the central perceptual mechanisms make no allowance for differences in conduction delays, the time of arrival at the cortex is not decisive for judgements of the time of occurrence of a stimulus.

Some justification is perhaps needed for assuming that the primary evoked potential in the somatosensory cortex is the event on which judgements about temporal and spatial patterns are likely to be based. There is much clinical and experimental evidence that fine discriminative sensation requires the integrity of the posterior column pathways of the spinal cord and the primary receiving area of the postcentral gyrus. No doubt it is only in this rather direct pathway, with only three synapses intervening between sense receptors and cortical neurones, that the necessary information on which to base resolution of the temporal and spatial features of the stimulus will be transmitted in a recognizable form.

Giblin (1960) has shown that the primary evoked potential is lost in patients with damage to the posterior columns and resulting loss of fine sensation, and we have confirmed this (Halliday and Wakefield, 1962). There are thus good grounds for associating this potential with the arrival of the relevant sensory information.

This must not be taken to imply that perception itself necessarily occurs early. While what is perceived may depend initially on the early discharge pattern of cortical neurones (and no doubt its subsequent analysis), there is evidence that this information does not necessarily become available to the subject's consciousness until later. This seems to follow conclusively from the evidence that a near-threshold stimulus can be rendered imperceptible by another stimulus coming up to 100 millisec. or more afterwards (Crawford, 1947; Battersby, Wagman, Karp and Bender, 1960; Halliday and Mingay, 1961).

The experiments of Piéron and Segal (1939) illustrate clearly the distinction between perception time and the perception of interval, and imply that these depend on separate central events. They found that it was possible to make two subthreshold stimuli perceptible when separated by an interval of up to 400 millisec. Since either stimulus alone was imperceptible, their perception must have depended on summation, but since both stimuli were perceived, separated by the appropriate interval, the temporal pattern of stimulation must have been preserved.

They made the important suggestion that, while the temporal pattern perceived was dependent on conduction along relatively direct pathways, the perceptual awareness of it depended on impulses arriving by slower multi-synaptic routes. These authors envisaged both these neuronal elements as being intra-cortical, but the principle they suggested finds support in the more recent work on the physiology of the slower polysynaptic brain-stem pathways. It is clear that while the specific evoked potential, travelling by the classical sensory pathway is unchanged in states of altered perceptual awareness (such as anaesthesia, concussion and hysterical loss of sensation), the volley ascending in the slower brain-stem pathways may be abolished. The latter therefore shows a much closer relationship with perceptual awareness, while the former appears to transmit the finer pattern on which the discriminative perceptions depend.

We are grateful to Professor P. Armitage for his advice, and in particular for suggesting to us the method of calculating the point of apparent simultaneity, and we should also like to thank Dr. E. A. Carmichael for his support and encouragement. Dr. R. Efron and Dr. G. Ettlinger read the paper for us in draft, and it has benefitted from their helpful criticisms.

## References

Battersby, W. S., Wagman, I. H., Karp, E., and Bender, M. B. (1960). Neural limitations of visual excitability: Alterations produced by cerebral lesions. *Arch. Neurol., Chicago*, **3**, 24–42.

Békésy, G. von (1957). Sensations on the skin similar to directional hearing, beats and harmonics of the ear. *J. acoust. Soc. Amer.*, **29**, 489–501.

Békésy, G. von (1959). Neural funneling along the skin and between the inner and outer hair cells of the cochlea. *J. acoust. Soc. Amer.*, **31**, 1236–49.

Crawford, B. H. (1947). Visual adaptation in relation to brief conditioning stimuli. *Proc. roy. Soc., B.*, **134**, 283–302.

Dawson, G. D. (1954). A summation technique for the detection of small evoked potentials. *Electroenceph. clin. Neurophysiol.*, **6**, 65–84.

Finney, D. J. (1952). *Statistical Method in Biological Assay*. London.

Giblin, D. R. (1960). The effect of lesions of the nervous system on cerebral responses to peripheral nerve stimulation. *Electroenceph. clin. Neurophysiol.*, **12**, 262.

Golla, F., and Antonovitch, S. (1933). The rate of conduction and refractory period of the human sensory neurone. *Brain*, **56**, 71–82.

Halliday, A. M., and Mingay, R. (1961). Retroactive raising of a sensory threshold by a contralateral stimulus. *Quart. J. exp. Psychol.*, **13**, 1–11.

Halliday, A. M., and Wakefield, G. S. (1962). Cerebral evoked responses in patients with dissociated sensory loss. *Electroenceph. clin. Neurophysiol.*, **14**, 786.

Hirsch, I. J., and Sherrick, C. E. (1961). Perceived order in different sense modalities. *J. exp. Psychol.*, **62**, 423–32.

Klemm, O. (1925). Uber die Wirksamkeit kleinster Zeitunterschiede im Gebiete des Tastsinns. *Arch. ges. Psychol.*, **50**, 205–20.

Ladefoged, P., and Broadbent, D. E. (1960). Perception of sequence in auditory events. *Quart. J. exp. Psychol.*, **12**, 162–70.

Piéron, H., and Segal, J. (1939). Sur un phénomene de facilitation rétroactive dans l'excitation électrique de branches nerveuses cutanées. *J. Neurophysiol.*, **2**, 178–91.

Pitman, J. R. (1958). A digital delay and time-scale generator. *J. Physiol.*, **142**, 3–5 P.

Titchener, E. B. (1910). *A Textbook of Psychology*. Part II. New York.

Wundt, W. (1903). Quoted by Woodrow, H., in chapter on "Time Perception" in Stevens, S. S. (1951). *Handbook of Experimental Psychology*. New York.

*Manuscript received 10th August, 1962.*

# EFFECTS OF STIMULUS PROBABILITY ON REACTION TIME IN A NUMBER-NAMING TASK

BY

GEORGE C. STONE and ENOCH CALLAWAY III

*From the Department of Psychiatry, University of California School of Medicine, San Francisco, and The Langley Porter Neuropsychiatric Institute*

Subjects were required to respond to the visual presentation of numerals by uttering syllables closely resembling the names of the numerals. Information in stimulus ensembles was varied by manipulating the number of alternative stimulus configurations that could appear, the relative frequencies of stimuli within an ensemble and the probability of a response being required. An increasing linear function was a good first approximation to the relation between reaction time and information transmission. Systematic deviations from this function were found and an attempt is made to explain them by introducing an intervening variable, "effective probability."

The proposition that dysjunctive reaction time is a linear increasing function of the amount of information transmitted from the stimulus to a response (Hick, 1952; Hyman, 1953; Crossman, 1953) has been shown to apply in an impressive variety of situations. Recently, this generalization has been challenged (Leonard, 1959; Mowbray and Rhoades, 1959; Seibel, 1959; Mowbray, 1960). These investigators present evidence that under conditions of optimum coding or of highly overlearned stimulus-response pairings the slope of the function is essentially zero.

We were interested in the possibility of studying the effects of manipulation of attention on the slope and intercept parameters of the function relating stimulus information and reaction time. We desired to present large amounts of information concerning the relative frequency of stimuli, without overtaxing our subjects by requiring an excessive number of responses. We therefore began some experiments in which a response was required on only a fraction of a total number of trials. The results of our preliminary experiments were surprising to us and before we were able to comprehend them thoroughly we were led to a reformulation of the question of the relation between stimulus probabilities and reaction time. On the basis of evidence to be described below it now appears to us that reaction time is not a function of the average amount of information required to be transmitted from stimulus to response, but rather of the logarithm of the effective probability of a stimulus weighted by the probability of a response to that stimulus. In the limiting case where effective probability is equal to objective probability and where every stimulus requires a measurable response, the formula for calculating estimated mean reaction time is identical with that for calculating information transmission. However, the new formulation leads to better handling of the partial response experiments described here and also permits the introduction of the concept of effective probability to account for deviations of specific reaction times from what would be predicted on the basis of objective probability.

## METHOD

The subjects, 8 inmates of a state penitentiary, were tested four days per week for a total of 16 or 20 sessions as described below. Each test session required 192 test responses in addition to 32 practice responses. The sessions lasted about one-half hour.

The subject's task was to make a verbal response to a visual display of a numeral presented on a Nixie tube, whenever a small neon bulb located one inch above the Nixie tube was lit. Each trial was initiated by the experimenter's pressing a switch that audibly activated a relay, reset an electronic timer to zero, transported the tape in a teletype reader into position for the selection of the stimulus and instituted a delay. This delay, which was approximately 1·5 sec., ended with the appearance of an Arabic numeral on a Nixie tube with or without the simultaneous illumination of the neon bulb. Under the 100 per cent. response condition, the neon bulb was lit each time the Nixie tube fired, but under the partial response condition it lit simultaneously with the Nixie tube only 40 per cent. of the time. (Selection of the value 40 per cent. was arbitrary.) The subject's instructions were that when the neon bulb was not lit, no response was required. When it was lit, he was to respond as quickly as possible with the syllable "bun" when he saw a one, "boo" for two, "bee" for three, and "bore" for four. The subject's vocal response was picked up by a microphone held about two inches from his lips by a harness arrangement. This response activated a voice key and interrupted the timing circuit. The reaction time was measured with a Hunter "Klock Counter" to the nearest millisec. from the instant the stimulus appeared until the voice key operated. After each trial the reading on the timer was recorded by the experimenter before the next trial was initiated.

*Experimental Design.* A total of eight conditions was used, varying in the amount of information presented by the stimulus display. Each of four values of $H_s$, the uncertainty as to which numeral would appear on the Nixie tube, was presented both under the 100 per cent. and partial response conditions. The eight conditions employed in each series are shown in Table I.

TABLE I

PROBABILITY OF OCCURRENCE OF VARIOUS STIMULUS-RESPONSE ALTERNATIVES IN EACH OF THE EIGHT EXPERIMENTAL CONDITIONS

| | | *Numeral on Nixie* | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | | 2 | | 3 | | 4 | | | | |
| *Pilot* | | *On* | *Off* | *On* | *Off* | *On* | *Off* | *On* | *Off* | $-p_s\log_2 p_s$ (*Formula* 1) | $-p_r\log_2 p_s$ (*Formula* 2) | $-p_r\log_2 p_e$ (*Formula* 3) |
| Condition | | | | | | | | | | | | |
| A | $p_s$ | 500* | 0 | 500 | 0 | 0 | 0 | 0 | 0 | 1·00 | 1·00 | 1·00 |
| | $p_r$ | 500 | 0 | 500 | 0 | 0 | 0 | 0 | 0 | | | |
| B | $p_s$ | 250 | 0 | 250 | 0 | 250 | 0 | 250 | 0 | 2·00 | 2·00 | 2·00 |
| | $p_r$ | 250 | 0 | 250 | 0 | 250 | 0 | 250 | 0 | | | |
| C | $p_s$ | 812 | 0 | 062 | 0 | 062 | 0 | 062 | 0 | 0·99 | 0·99 | 0·68 |
| | $p_r$ | 812 | 0 | 062 | 0 | 062 | 0 | 062 | 0 | | | |
| D | $p_s$ | 500 | 0 | 250 | 0 | 125 | 0 | 225 | 0 | 1·75 | 1·75 | 1·36 |
| | $p_r$ | 500 | 0 | 250 | 0 | 125 | 0 | 125 | 0 | | | |
| E | $p_s$ | 202 | 298 | 202 | 298 | 0 | 0 | 0 | 0 | 2·30 | 1·97 | 1·90 |
| | $p_r$ | 500 | 0 | 500 | 0 | 0 | 0 | 0 | 0 | | | |
| F | $p_s$ | 101 | 149 | 101 | 149 | 101 | 149 | 101 | 149 | 3·30 | 2·97 | 2·89 |
| | $p_r$ | 250 | 0 | 250 | 0 | 250 | 0 | 250 | 0 | | | |
| G | $p_s$ | 329 | 483 | 025 | 037 | 024 | 037 | 025 | 037 | 2·30 | 1·96 | 1·82 |
| | $p_r$ | 812 | 0 | 062 | 0 | 062 | 0 | 062 | 0 | | | |
| H | $p_s$ | 202 | 298 | 101 | 149 | 051 | 074 | 051 | 074 | 3·05 | 2·72 | 2·46 |
| | $p_r$ | 500 | 0 | 250 | 0 | 125 | 0 | 125 | 0 | | | |

* Probability of Occurrence × $10^3$.

Each subject was tested on two of the eight conditions per testing session. Four testing sessions were required to expose each subject to every condition. These four test sessions constituted one complete replication of the experiment. Within each replication the order of presentation of the conditions to subjects was randomized. In subsequent replications new random assignments of subjects to orders were made.

Six subjects completed five replications and two subjects, four replications. The total testing period for each subject lasted from two to three months.

On every test day a subject was given 16 trials of practice on the first condition scheduled for that day, followed by testing on half of the trials for that condition. Then 16 trials of practice were given on the second condition to be tested on that day, followed by testing on half the trials for the second condition. After a 3-min. rest the second half of the second condition was run, followed by the second half of the first condition. In addition to practice on each condition, subjects were also orally informed beforehand of the probabilities of each numeral's appearance. One condition each day was from the 100 per cent. response group and one from the partial response group. In the 100 per cent. response group a complete set of trials was 128; in the 40 per cent. response series a complete set was 64 responses. (148 numerals were flashed on the Nixie tube but the pilot light lit only 64 times).

Motivation for fast and accurate performance was provided through a system of bonuses. The subjects were paid on a sliding scale on the basis of the overall speed of their reaction times throughout the four sessions that constituted a complete replication. Each week the eight subjects were divided into four pairs on the basis of corrected speed and were paid four different fees—$1.00, $1.50, $2.00 and $2.50; the maximum pay was $10.00 per month. This amount was expected to produce a satisfactory level of motivation, since pay to inmates at this prison is usually limited to $6.00 per month. The time scores were corrected for pay purposes as follows: the times for each response were totalled and a penalty of 2 sec. for each error was added to the total score. Errors consisted of incorrect naming responses, of responses in advance of a signal or of other responses that showed the subject was not "playing according to the rules."

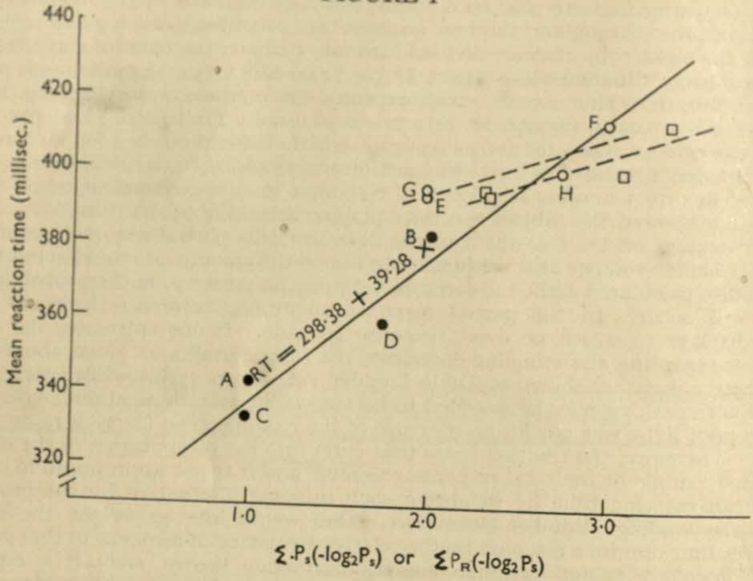
Some comment on the treatment of the data from the partial response conditions is necessary. Under 100 per cent. response conditions the amount of stimulus information can be readily calculated from the formula $-\Sigma p_s \log_2 p_s$, where $p_s$ is the probability that stimulus will occur. In the partial response condition, however, there are several stimulus displays to which no overt response is made. If one calculates the average uncertainty regarding the stimulus ensemble, the same formula as given above can be used. These values are shown in Table I under $-\Sigma p_s \log_2 p_s$. But while reactions that consist of not speaking may be assumed to be associated with those stimulus displays in which the pilot light was not lit, no measure of the reaction time for such responses was obtained. Therefore, the reaction times that enter into the computation of the mean are but a biased sample of the total response ensemble and it is not appropriate to estimate overall mean reaction time by weighting each response alternative by the probability of the corresponding stimulus alternative. One would like to weight the stimulus alternatives that require a response by the relative frequency of response to that stimulus. But this procedure is not legitimate under information theory, since it is essentially meaningless to speak of the information content of a component of an ensemble of possible messages. We therefore employed the alternative formulation of predicting reaction time only from the negative logarithm of the probability of a particular stimulus weighted by the probability of a response to that stimulus. The equation here is $-\Sigma p_r \log_2 p_s$ and the corresponding values for each condition are given in Table I.

## RESULTS

Mean reaction times for all subjects on all replications for each condition are shown in Figure 1. A regression line has been fitted to the means of conditions A through D as shown. The linear component of this regression is significantly different from zero with a probability less than 0·001. The means for conditions E through H are plotted in two different places: (1) On the left (the circles), under the assumption that mean reaction time is a function of the average uncertainty regarding the stimulus ensemble (Formula 1, Table I), and (2) on the right (the squares), under the assumption that each reaction time is a function of the log of the probability that a given stimulus will occur. Mean reaction time will then be predicted by weighting the log probability of each possible stimulus by the probability that a response to it can be measured (Formula 2). Regression lines fitted to these four points have significantly shallower slopes than the regression line fitted to the means for conditions A through D. The sum of the squared deviations of the right-hand set of points from the regression line, $RT = 298 + 39X$ is substantially, though not significantly, smaller than that for the set of points based on stimulus uncertainty. The same may be said of the points based on each of the five replications or on each of the eight subjects independently. In our further analysis we therefore based our expectations on the formulation that reaction time was a function of the log of th probability of the stimulus reacted to. It thus became possible to look at mea reaction times to individual stimuli.

The deviations of the eight conditions from regression that can be seen i Figure 1 were highly significant ($p < 0·001$). Furthermore, the pattern of deviations

FIGURE 1



was consistent from replication to replication and from subject to subject. For example, the depression of the mean for condition D below the overall regression line occurred in 12 of the 13 separate sets of means. Summary of these deviations can be found in Table II.

To pursue the question of the significance of these deviations, the mean reaction times for individual stimuli were plotted as a function of their probability of occurrence. These values are shown in Figure 2. The regression line shown is that derived from the overall regression $RT = 309 + 33X$, where X is log 1/p and p is the probability that the numeral will appear together with the neon light indicating that a response is necessary. Each data point is represented by two symbols, the letter indicating the experimental condition and the number of the stimulus numeral. Thus, the point H-1 stands for the mean reaction time obtained in condition H to the numeral 1.

As Hyman pointed out in his original article (1953, p. 194), "The components within a condition interacted with each other in such a way that the reaction time to a low information component was high and the reaction time to the high information component was markedly lower than would be predicted on the basis of the regression line fitted to the means of the conditions." With the semantic modification of describing the differences between points in terms of their probability rather than their information content, our data are in thorough agreement with the findings indicated by Hyman. That is, high probability stimuli elicited somewhat slower responses than one would expect, while low probability stimuli elicited somewhat faster responses than one would expect. In particular, the deviations of the points from the overall regression line seemed to conform very well to the rule that reaction times would be slow when the probability of a stimulus was greater than 1/n (n = number of alternatives in the ensemble) and fast when the probability of the stimulus was less than 1/n. Four stimuli occurred with probability greater than 1/n (C-1, D-1, G-1 and H-1). All of these fell to the right of the regression line. All points that are markedly to the left of the regression line, on the other hand, represent stimulus displays having probability substantially lower than 1/n.

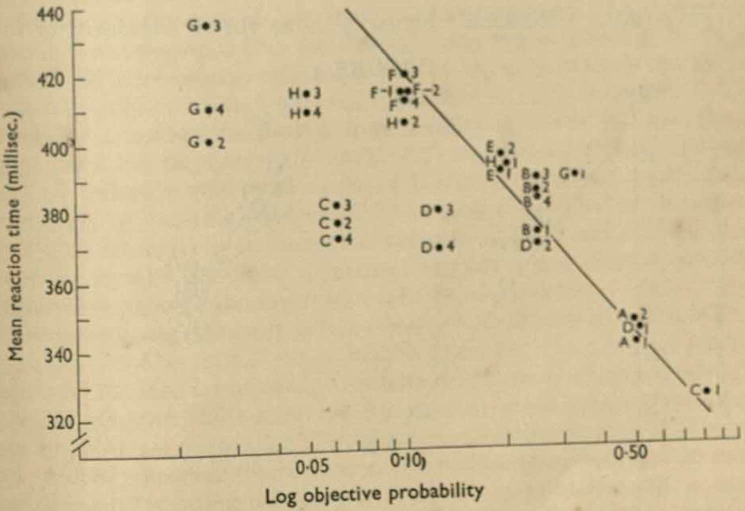
TABLE II

DIRECTION OF DEVIATIONS OF MEANS FROM OVERALL REGRESSIONS

| | | *Overall regression* | *Condition*[*] 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| Replication | 1 | RT = 330 + 36x | + | + | − | − | + | − | + | + |
| | 2 | 306 + 35x | + | + | − | − | + | − | + | − |
| | 3 | 304 + 32x | − | + | + | − | + | + | + | − |
| | 4 | 297 + 32x | o | + | − | − | + | − | + | − |
| | 5 | 300 + 29x | + | + | − | − | + | − | + | − |
| Subject | 1 | 330 + 34x | + | + | − | − | − | − | − | o |
| | 2 | 330 + 33x | − | + | − | − | + | − | + | − |
| | 3 | 332 + 30x | + | + | − | − | + | − | + | − |
| | 4 | 322 + 33x | + | + | − | − | + | + | + | − |
| | 5 | 309 + 38x | + | + | − | + | − | + | o | − |
| | 6 | 293 + 42x | + | + | − | − | + | + | o | − |
| | 7 | 263 + 32x | − | − | − | − | + | − | + | − |
| | 8 | 279 + 24x | + | − | − | − | − | − | + | o |

[*] +, Mean RT greater than predicted; o, within 1 millisec. of prediction; −, Mean RT less than predicted from regression.

FIGURE 2



We considered it possible to find a sensible mathematical function to describe the observed relations between objective probability and mean reaction time. To this end we defined an intervening variable, "effective probability" ($p_e$), that would be a function of objective probability ($p_o$), the number of events in an ensemble (n), and the probability of the most likely event ($p_m$). We further required that mean reaction time should be a linear function of the logarithm of effective probability. In short, we defined $p_e$ with two equations:

$$p_e = F(p_o, n, p_m)$$
$$RT = a + b \log p_e$$

The term "effective probability" is preferred because of the established use of "subjective probability" in utility theory (*vide* Adams, 1960). While the two concepts have heuristic similarities, it is unlikely that they are identical, and it would be well to distinguish them until the relationships can be explored. (Since this paper was completed we have found a similar suggestion made by S. Kornblum in "Reaction Time to Sequential Stimulus Presentations." Ph.D. thesis, Univ. of Michigan, 1959.)

We made the working assumption that $p_o = p_e$ when all stimuli appear with frequency 1/n, and were thereby able to graph $p_e$ against $p_o$. Study of the resulting sigmoid graph suggested two constraints in the selection of the function: (1) it appeared to have a lower symptote of 1/n; (2) $p_m$ seemed to influence the point of inflection, which occurred in the interval when $p_o$ was between 1/n and $p_m$. Following these guidelines, we found several functions that gave moderately good fits to the data points of Figure 2. Best was

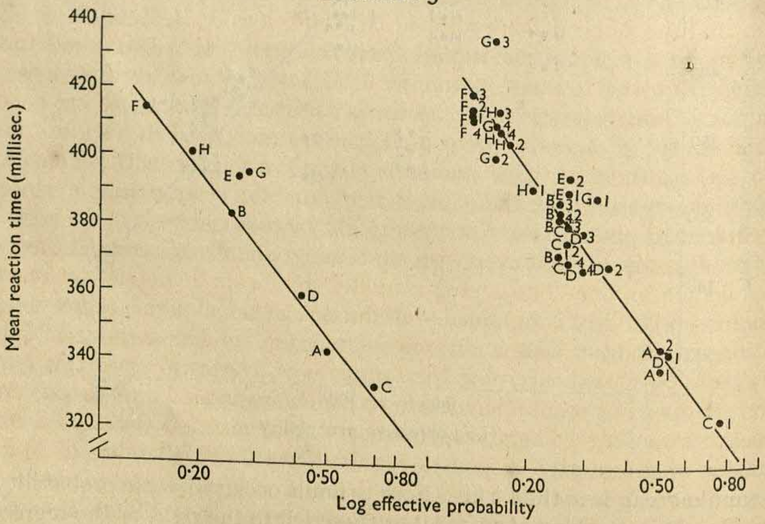
$$p_e = 1/n + \frac{(p_m - 1/n)}{1 + 10^{Kp_a}}$$

where $p_a$ is

$$\frac{p_m + 1/n}{2} - p_o$$

and K is an arbitrary constant. Figure 3 shows the fit obtained for individual points and for condition means using this function and setting $K = 3{\cdot}5$. While the data do not fit perfectly, Figure 3 is a clear advance over Figure 2. Figure 3 indicates

FIGURE 3



the utility of effective probability for predicting reaction times to particular stimuli. On the other hand, when deviations of the condition means, rather than individual stimulus means, from their regression lines were compared no significant improvement was found to result from the shift from $p_o$ to $p_e$. However, the points based on $p_e$ fit very closely on a straight line for all conditions except E and G. If these two means are excluded, the F ratio for improved fit of the $p_e$ regression relative to the $p_o$ regression is 11·3, significant at the 0·02 level. Some reasons for the excessive deviations of the E and G condition means are offered below.

## DISCUSSION

It would appear that to a reasonably good approximation reaction time to a stimulus is a function of the number of alternative stimuli that can occur. With practice one can make use of the fact that a stimulus is particularly common, but one rarely deals with stimuli as though they were much less common than if all stimuli in the ensemble were equally probable. However, examination of Figure 3 illustrates that other factors are operating.

One such factor must be at work in conditions E and G. These conditions had in common stimuli with high probabilities of occurring when no response was required. In condition G the numeral "1" appeared a total of 52 times in each replication, but was responded to only 21 times. On 31 occasions any tendency to respond had to be inhibited. It appears that excessive practice of the response inhibition offset the gain that would otherwise have accrued from the high probability of occurrence of this stimulus. Similarly, both numerals in condition E were presented many times under nonresponse conditions and their mean reaction times are somewhat slowed. Why the effect is not seen also in conditions F and H is not clear.

One more factor leading to deviations from the regression lines is the consistent differences between reaction times to the various stimuli within an ensemble even when probabilities of occurrence are equal. In all conditions of a single replication the numeral 1 was responded to 396 times; 2, 204 times; 3 and 4, 84 times each. Responses to "1" were consistently faster than to "2" even in those conditions where their probabilities of occurrence were equal (conditions A, B, E, and F ). Thus, one might think of a residual facilitation from conditions where "1" had the greatest frequency. But the response to numeral "2" was almost always slower than to numeral "4," though it was made 2½ times as often. Also, in all six conditions in which numerals "3" and "4" were used, the reaction time to "4" was faster than that to "3," although these numerals always occurred with equal frequency. Therefore, the reasons for these differences are not clear. Simple motor factors (time required to phonate the response) might be reasonable.

Our results indicate that even with so seemingly natural or compatible a stimulus-response coding as one-"bun," etc., a highly significant regression of reaction time on stimulus probability is obtained. Reduction in slope of this regression from one replication to another, although not significant, offers suggestive support for Mowbray's hypothesis concerning the influence of overlearning on this relationship. However, it does not seem appropriate to conclude, as Mowbray (1960) does, "that the generalization regarding the increase of choice reaction times as a function of the number of alternative responses required has to be abandoned" (p. 200). It would seem more appropriate to consider that the nervous system processes relatively novel S-R pairs via different means than overlearned ones. Such a suggestion has been made by Davis, Moray and Treisman (1961) in their argument for progressive "automatization" of responses with extended training.

The suggestion has been made that in our set of responses it would be possible to prephonate the "b . . ." and thus shorten the reaction time. If the procedure were followed by the subjects, there might be a tendency to equalize RTs under the various conditions; since when a time-consuming decision had to be made, the "b . . ." could simply be prolonged. We cannot entirely rule out the possibility that such behaviour may have occurred, although we instructed against it. Further, when one of the authors served as subject in preliminary experiments, he found that this technique actually increased his reaction time. In any case, the operation of this factor would be against the findings that we report of differences between reaction time under various conditions.

Discussion of absolute values of reaction times is of little use, except to validate the general method, as there are too many factors that can influence them. Means for our 2-bit condition ranged from 325 to 429 millisec. for our eight subjects. In somewhat comparable experiments using visual presentation of stimuli Hyman (1953) found means ranging from about 350 to 600 millisec.; Mowbray (1960), means of 459 millisec.; Suci *et al*, over 650 millisec. Thus, our reaction times are in the same range as other similar experiments have found, but somewhat faster than most. Davis, Moray and Treisman (1961), using auditory presentation of stimuli and an "imitative response," found substantially faster reaction times, of the order of 200 millisec.

Comparison of slopes, or rate of gain of information among experiments, may be more meaningful since constant errors are cancelled. Considering the same set of experiments we find the rate of gain of information was low, about 5·6 and 5·8 bits/sec., in the experiments by Suci (Young group) and Hyman, where very unpracticed S-R pairs were used. In Mowbray's Experiment II and in our experiment, where more practiced or "compatible" S-R pairs were employed, rates of 25 and 30 bits/sec. were found, while in the experiment by Davis *et al*, using an imitative response of possibly maximum compatibility, the rate was between 65 and 70 bits/sec. Mowbray's S-R pairs in the experiment can be considered "1"—Silence; "2"—Silence; . . . . "8," eight. Apparently the degree of practice of the entire S-R ensemble is the crucial factor. Since one dislikes to conceive of an infinite rate of gain of information, it seems preferable to suppose that results such as those of Mowbray in his Experiment I and of Leonard (1959) and others who find zero slope are due to the by-passing of the attention mechanisms of information processing and the substitution of some reflex process as Davis *et al* have suggested.

The particular function chosen to evaluate effective probability in this report is, of course, *ad hoc*, although some rational constraints were imposed. However, we feel that the concept of effective probability will provide a substantial improvement in the attempts to deal quantitatively with reaction times to the several elements of a relatively novel stimulus ensemble. We have not attempted to apply our formula to other data, except for the illustrative case given by Hyman (1953, p. 194) in his Experiment II, Conditions 3 for the frequent and infrequent stimuli:

| | | |
|---|---|---|
| Hyman predicted, using objective probability .. | 258 | 824 |
| He observed .. .. .. .. .. .. | 306 | 585 |
| We predicted, using effective probability .. .. | 304 | 553 |

This gain in precision was achieved using the same value for K (3·5) as was used in analysis of our data. The solution appears to have some generality, therefore. Other functions can also fit the data. For example

$$p_e = p_o + 1/ne^{-Kp_o} - (1 - p_m)e^{-K(1 - p_o)}$$

is nearly as good as the function chosen. Ultimately, of course, determination of the function must depend upon its articulation with other bodies of theory in the manner attempted by Luce (1960) and Rapaport (1959).

Collaboration and assistance from Mr. Leon Tabory, Mr. Luther Thomas, and the California Department of Corrections made this study feasible. The authors are also indebted to Mrs. Patsy Bodle for technical assistance. The work was supported by pilot study funds of the California Department of Mental Hygiene and the Office of Naval Research, NONR 2931 (00), project NR 144134. Reproduction in whole or in part is permitted for any purpose of the United States Government.

## REFERENCES

ADAMS, E. W. (1960). Survey of Bernoullian Utility Theory. In: SOLOMON, H. (Ed.), *Mathematical Thinking in the Measurement of Behavior*. Glencoe, Illinois.

CROSSMAN, E. R. F. W. (1953). Entropy and choice time: The effect of frequency unbalance on choice response. *Quart. J. exp. Psychol.*, **5**, 41–51.

DAVIS, R., MORAY, N., and TREISMAN, A. (1961). Imitative responses and the rate of gain of information. *Quart. J. exp. Psychol.*, **13**, 78–89.

HICK, W. E. (1952). On the rate of gain of information. *Quart. J. exp. Psychol.*, **4**, 11–26.

HYMAN, R. (1953). Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, **45**, 188–96.

LEONARD, J. (1959). Tactual choice reactions: I. *Quart. J. exp. Psychol.*, **11**, 76–83.

LUCE, R. D. (1960). Response latencies and probabilities. In: ARROW, K. J., KARLIN, S., and SUPPES, P. (Eds.). *Mathematical Methods in the Social Sciences*, Proceedings of the First Stanford Symposium, Stanford University Press.

MOWBRAY, G. H. (1960). Choice reaction times for skilled responses. *Quart. J. exp. Psychol.*, **12**, 193–202.

MOWBRAY, G. H., and RHOADES, M. V. (1959). On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, **11**, 16–23.

RAPAPORT, A. (1959). A study of disjunctive reaction times. *Behav. Sci.*, **4**, 299–315.

SEIBEL, R. (1959). Discrimination reaction time as a function of the number of alternatives and of the particular stimulus-response patterns. *Amer. Psychol.*, **14**, 396.

SUCI, G. J., DAVIDOFF, M. D., and SURWILLO, W. W. (1960). Reaction time as a function of stimulus information and age. *J. exp. Psychol.*, **60**, 242–4.

*Manuscript received 4th September, 1962.*

# SHORTER ARTICLES AND NOTES

## THE INFLUENCE OF SIGNAL PROBABILITY AND THE NUMBER OF NON-SIGNAL CATEGORIES ON SELECTIVE RESPONSE TIMES

BY

JOHN BREBNER and IAN GORDON

*From the Psychological Laboratory, University of Exeter**

Signal probability and the number of non-signal categories were varied independently in a selective response task. Response latencies were significantly affected by changes in signal probability but not by changes in the number of non-signal categories.

### INTRODUCTION

Mowbray (1960) in an investigation into choice reaction times using the overlearned response of naming Arabic numerals, found that when each stimulus item had a separate response associated with it, increasing the number of stimulus categories did not result in an increase in response latencies. In a second condition in which subjects responded only to one of the stimulus categories, however, increasing the number of categories did produce longer response latencies. Mowbray suggested that the slowing observed with the larger ensembles in the latter condition (which we shall refer to as a "selective response task") might be due to the fact that the ratio of signal to non-signal digits decreased as the number of alternatives was raised.

As was noted earlier (1962), however, increasing the number of alternatives in a selective response task alters not one but several variables, e.g. the proportion of signal to non-signal items, and the number of non-signal categories. Selective response tasks provide a convenient situation in which to study these variables independently, and the present experiment aimed at investigating the effects upon selective response latency of signal probability and the number of non-signal categories.

### APPARATUS

The apparatus was the same as that previously used by the authors (1962). A system of two uniselectors produced a sequence of digits which repeated only after 2,450 presentations. Subjects sat at a table in a dimly lit room, facing and one metre from a digital display unit on which illuminated Arabic numerals 1 in. in height appeared at regular time intervals. The apparatus was driven by an electronic metronome set at a regular rate, the beat of which was audible to subjects so that it was possible to anticipate when the next digit was due to appear. This kept the subject's uncertainty about the time of occurrence of digits at a low level. In all conditions digits were illuminated for 650 millisec.

One new feature was that the response required in the present experiment was to operate a microswitch with the forefinger of the dominant hand. Closing this switch stopped a chronotron (1,000 millisec. Full Scale Deflection) which started on the appearance of the digit "2."

### METHOD

Four main conditions were used in the present experiment, in all of which responses were required only to the digit "2." Signal probability and the number of non-signal digits were varied independently as shown in Table I. In the 2-digit condition digits "1" and "2" only were used, while in the 4-digit condition digits "1"–"4" were presented to the subject. Digits were presented at rates of 60 per min. when the signal probability was 0·25, and 24 per min. when the signal probability was 0·625. The average rate of appearance of the signal digit ("2") was, therefore, 1 per 4 sec. in all conditions.

---

* Both authors now at the University of St. Andrews, Queen's College, Dundee.

Four male undergraduates took part in the experiment and a Latin Square design was used in an attempt to ensure that the order in which subjects performed the conditions did not materially affect the results. At the start of each test session a short practice period was given which lasted until 25 responses had been made by the subject. In the actual test sessions 200 reaction times were obtained from each subject. Each session lasted approximately 8 min. but a 3-min. rest pause was introduced after the first 100 responses had been obtained. Before the start of each test session subjects were informed which digits would be used in that particular condition and warned of their relative frequencies and the rate at which the digits would be presented. Subjects were instructed to respond as rapidly as possible to the appearance of the signal digit, to respond to no other digit and to refrain from making errors. This last instruction appears to have been reasonably successful in that no subject made more than 7 errors in any test session.

The authors' decision to hold signal rate rather than digit rate constant while providing a constant mean intersignal interval, does not exclude the possibility that the rate of presentation of digits may itself affect selective response latencies. Accordingly, two further conditions (*c*) 2(0·625) at a digit rate of 60 per min., and (*f*) 2(0·25) at a digit rate of 24 per min., were included in the experiment, *v* Table I. These two further conditions were carried out when subjects had completed their initial four test sessions, two of the subjects being tested in the order of conditions *c*–*f* and the other two subjects in the order *f*–*c*. 100 RTs were recorded from each subject in both these additional conditions.

## RESULTS AND DISCUSSION

Table I shows the subjects' mean selective response times for the six tests conditions. Conditions (*c*) and (*f*) are the additional conditions referred to in the section under Method. The distribution of response times in the four main conditions of the experiment are shown in Figure 1.

TABLE I

MEAN SELECTIVE RESPONSE TIMES (MILLISEC.) FOR EACH SUBJECT IN THE SIX CONDITIONS*

| | *Digit rate* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 60 *per min.* | | | | | | 24 *per min.* | | | | | |
| | (*a*) 2(0·25) | | (*b*) 4(0·25) | | (*c*) 2(0·265) | | (*d*) 2(0·625) | | (*e*) 4(0·625) | | (*f*) 2(0·25) | |
| Subject | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| J.B. | 376 | 38·1 | 410 | 37·9 | 336 | 34·3 | 302 | 39·2 | 341 | 42·0 | 358 | 45·6 |
| A.G. | 374 | 20·9 | 386 | 39·5 | 307 | 37·6 | 282 | 27·0 | 298 | 37·3 | 317 | 22·7 |
| D.S. | 360 | 54·6 | 367 | 51·3 | 337 | 37·5 | 302 | 42·6 | 311 | 42·4 | 339 | 41·9 |
| S.B. | 393 | 46·7 | 401 | 52·4 | 304 | 50·1 | 304 | 56·3 | 305 | 51·0 | 322 | 60·8 |
| Mean | 376 | | 391 | | 321 | | 298 | | 314 | | 334 | |

* Numerals outside brackets refer to the number of alternative digits used in that condition, bracketed numbers refer to the signal probability.

The results of *t*-tests carried out between the various conditions are shown in Table II.

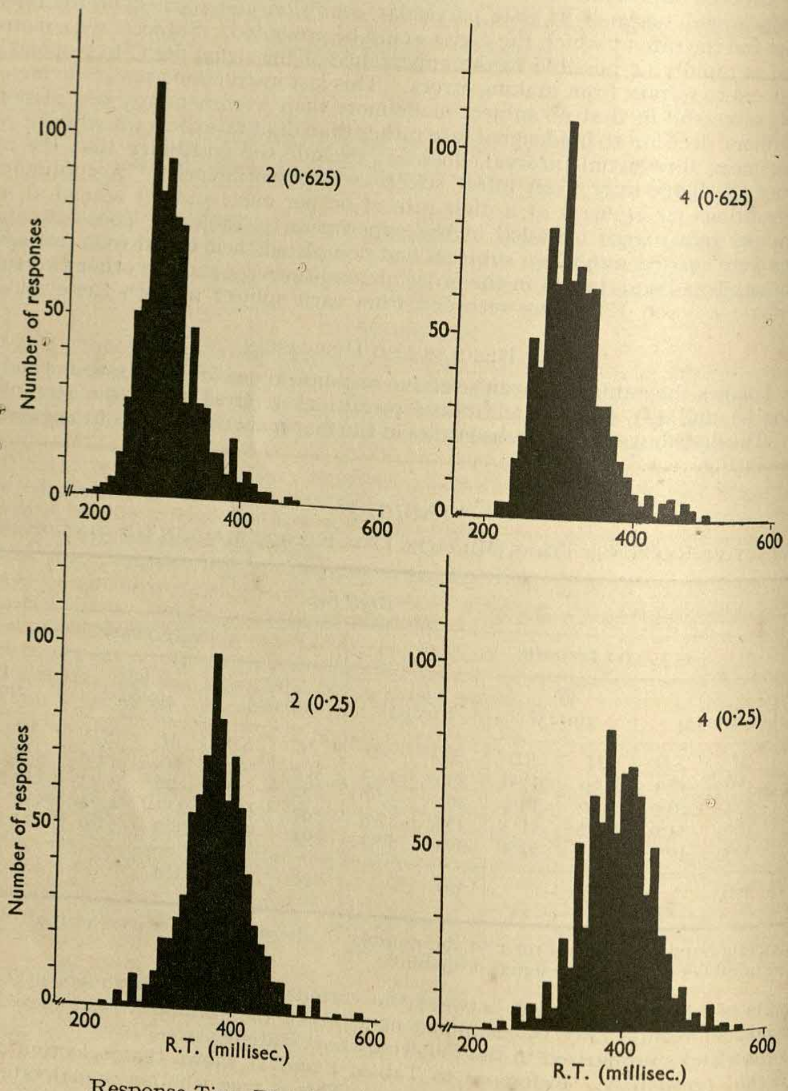
The present results may be examined in terms of the variables manipulated in the investigation and summarized in the following short statements:—

(i) *Signal probability*. Reference to Tables I and II shows that selective response times shorten as signal probability increases. This is true whether signal rate is held constant or digit rate is held constant, and the ensemble size is two or four digits. In all these cases increasing the signal probability from 0·25 to 0·625 results in a significant shortening of mean selective response times.

(ii) *Non-signal categories*. The present results indicate that increasing the number of non-signal categories serves to lengthen selective response times only slightly. However, it is the case that for all subjects mean response times are, to a variable extent, affected by the change in the ensemble of digits. It may also be noted that the difference in mean selective response times between the two levels of signal probability (with a constant signal rate) is very similar whether two or four digits are used.

(iii) *Digit rate.* From Table I it may be seen that the effect of increasing the rat which digits are presented, for a given signal probability, is to increase mean selecti response times. This effect of digit rate appears to be more marked when sig probability is low (*v* columns (*a*) and (*f*) Table I) than when signal probability is hi

FIGURE I



Response Time Distribution in the Four Main Conditions

(*v* columns (*c*) and (*d*) Table I), but in neither case is the 5 per cent. level of significance attained. The alteration of mean selective response times associated with changes in digit rate may be due to the shorter mean intersignal intervals which result from increasing the digit rate while holding signal probability constant. The existence of an inverse relationship between selective response times and intersignal interval has been demonstrated by Mowbray (1960). The larger difference in selective response times between the two rates of presentation which occurs with the lower signal probability and just falls short of the 5 per cent. level of significance, could on this view be due to the relatively greater difference between the mean intersignal intervals (one per 4 sec.—one per 10 sec.) in conditions (*a*) and (*f*) compared with that between conditions (*c*) and (*d*) (one per 1·6 sec.—one per 4 sec.).

TABLE II

*t*-VALUES FROM THE MEAN DIFFERENCES BETWEEN CONDITIONS, AND ASSOCIATED PROBABILITIES

| *Condition* .. .. | *a* | *b* | *c* | *d* | *e* | *f* | *Mean difference in millisec.* | *t* | *p* |
|---|---|---|---|---|---|---|---|---|---|
| *Categories and Signal Probability* .. | 2(0·25) | 4(0·25) | 2(0·625) | 2(0·625) | 4(0·625) | 2(0·25) | | | |
| *Digit rate* .. .. | 60 per min. | 60 per min. | 60 per min. | 24 per min. | 24 per min. | 24 per min. | | | |
| | ← | — | — | → | | | 78 | 10·02 | <0·01** |
| | | ← | — | — | → | | 77 | 8·50 | <0·01** |
| | | | | ← | — | → | 36 | 4·69 | <0·02* |
| | ← | — | → | | | | 55 | 3·75 | <0·05* |
| | ← | — | — | — | — | → | 42 | 3·17 | <0·10 |
| | | | ← | → | | | 23 | 2·88 | <0·10 |
| | ← | → | | | | | 15 | 2·40 | <0·10 |
| | | | | ← | → | | 16 | 1·99 | <0·20 |

The present findings may have some implications for the interpretation of results obtained in choice response situations in which each signal is associated with a particular response. In choice response situations, increasing the size of the signal ensemble results in a lengthening of response latencies except when the subject is very highly practised (Mowbray and Rhoades, 1959), or the response is an overlearned one, such as the verbal response of naming digits (Mowbray, 1960), or when each signal is presented directly to the responding member (Leonard, 1959). When ensemble size is increased and, as is usually the case, all items are presented with equal frequency, the stimulus situation is changed in at least two ways. Firstly, more separate categories requiring identification are presented to the subject and, secondly, the probability that a particular response will be required is reduced. The slowing observed with larger ensembles in choice response situations could, then, be due either to reduced signal probability or to an increase in the number of categories to be identified, or to an interaction of both effects. In this connection Fraisse and Blancheteau (1962) have found that the threshold for recognition increases as a log function of the number of alternatives in the ensemble, and it is interesting to note that Hick (1952) showed a relationship between the information (in the technical sense of the negative logarithm of the probability) of a choice response condition and the reaction times obtained in that condition, but offered as a tentative conceptual model a process of progressive dichotomization among the possible alternatives in the situation.

From the experimental situation described in this paper, in which signal probability and the number of categories presented have been manipulated separately, it would appear that the effect of signal probability, at least with the small increase of categories studied here, is the factor exerting the major influence on response latency. This raises the difficulty, however, that selective response times increase as signal probability decreases, whereas choice response times do not vary with signal probability when a highly practised response is used. This problem might be resolved by considering that the operative variable is not the probability of signals but rather the ratio of signals to non-signals which Mowbray has already stated he believes to be important (1962). However, it must also be at least a possibility that the differences in latencies observed in selective response tasks would disappear with prolonged practice, since although one response studied so far—that of naming Arabic numerals—is undoubtedly overlearned, that of naming selectively may not be. The fact of having to withhold responses to all but one of the display items may, in the early stages of practice, serve to increase the time taken to respond to the signal item in proportion to the number of non-signal items presented.

The authors are indebted to Drs. A. C. Staniland and J. Szafran for their advice and criticisms throughout this research, and gratefully acknowledge the financial assistance afforded them by the Nuffield Foundation Trust and the Department of Scientific and Industrial Research.

## REFERENCES

BREBNER, J., and GORDON, I. E. (1962). Selective response times with a constant signal rate. *Quart. J. exp. Psychol.*, **14**, 113–6.

FRAISSE, P., and BLANCHETEAU, M. (1962). The influence of the number of alternatives on the perceptual recognition threshold. *Quart. J. exp. Psychol.*, **14**, 52–5.

HICK, W. E. (1952). On the rate of gain of information. *Quart. J. exp. Psychol.*, **4**, 11–26.

LEONARD, J. A. (1959). Tactual choice reactions. *Quart. J. exp. Psychol.*, **11**, 76–83.

MOWBRAY, G. H., and RHOADES, M. V. (1959). On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, **11**, 16–23.

MOWBRAY, G. H. (1960). Choice reaction times for skilled responses. *Quart. J. exp. Psychol.*, **12**, 193–202.

MOWBRAY, G. H. (1962). Some remarks concerning Brebner and Gordon's Paper "Ensemble size and selective response times with a constant signal rate." *Quart. J. exp. Psychol.*, **14**, 117–8.

*Manuscript received 16th July, 1962*

# DIRECTIONAL INFORMATION AND APPARENT MOVEMENT—THE APPARENT MOVEMENT TO APPARENT SIMULTANEITY THRESHOLD

BY

M. A. JEEVES

*From the Department of Psychology, University of Adelaide*

This paper reports a study of changes in apparent movement to apparent simultaneity thresholds as a function of the directional information of the second point of light. An earlier paper (Jeeves and Bruner, 1956) reported changes in thresholds from apparent movement to apparent successiveness when the same experimental variable was manipulated. In interpreting the results the authors proposed an hypothesis in terms of "attentional disarticulation." Brown (1956) commenting on this earlier paper put forward an alternative explanation of the results in conventional informational terms and he predicted the changes in the apparent movement to apparent simultaneity thresholds which should occur if studied under the same experimental conditions. He further showed that his predictions would differ from predictions made on the basis of our own explanation. The results of the experiment reported here uphold the predictions of our own earlier theory and fail to fit those made by Brown. However, in attempting to understand the results from the two experiments considered together, it is suggested that a factor must be included which was omitted from our own earlier explanation but emphasized by Brown. This factor as stated by Brown (1956) is "that the threshold for a discrimination tends to increase with the information content of any discrimination which has to be made concurrently."

## INTRODUCTION

In an earlier paper (Jeeves and Bruner, 1956) we showed that none of the studies of apparent movement reported in the literature up to that time had taken account of the directional information in the movement perceived. We suggested that the greater the alternative number of directions that movement can take in a given situation, the greater the directional information (using information in the Shannon sense) the movement may be said to contain. Our hypothesis at the outset of the previous study had been that the larger the number of alternative directions that apparent movement could take, the poorer would be the apparent movement, but in fact the results obtained were quite opposite.

In attempting to understand these earlier findings we suggested an explanation in terms of "attentional disarticulation" of the perceptual field under instructions to expect movement in any of eight directions (as compared with either of two directions). In amplifying the meaning of "attentional disarticulation" we suggested that to distribute one's attention over eight possible pathways of apparent movement may be tantamount to reducing the kinds of constraints that operate in a field to keep events and points in that field segregated from each other. In a note later the same year Brown (1956) expressed dissatisfaction with our explanation and at the same time advanced his own alternative explanation. Detailed comment upon Brown's note will be reserved for the discussion at the end of this paper but part of it is of direct relevance at this point since Brown showed how his explanation and ours should make opposite predictions as to the outcome of the present study. Thus Brown predicted that "with increase in the directional information of the second stimulus, both the transition between apparent simultaneity and apparent movement and the transition between apparent movement and succession should occur at higher values of $t$, the interval between the two stimuli." According to Brown's understanding of our earlier theory he wrote "in relation to the simultaneity/apparent movement transition, expectancy for movement will presumably operate on descending trials. Hence . . . directional information will lower the value of $t$ for the transition on descending trials and will not affect it on ascending trials."

We shall return to a consideration of these two theories and the predictions they make after the experiment has been described in detail and the results set out.

## Procedure

Thirty-two female first-year students were used as subjects. The stimuli for apparent movement were two pinpoint spots of fluorescent white light in a black unstructured field in a high-speed electronic tachistoscope. The two pinpoints of light were 0·37 in. apart and 17 in. from the eyepiece of the subject. The lighting used in each half of the tachistoscope was two mercury-argon discharge tubes placed behind an aperture approximately 0·8 mm. in diameter. The spots of light were seen in a visual field either through or on a half-silvered mirror, the centre point of which was 31 cm. from the centre of the eyepiece. The half-silvered mirror had equal transmission and reflectance characteristics so that the two points of light were of equal apparent intensity.

In the presentation of the movement stimuli the first light appeared in the centre of the black field for 1·0 sec.—a length of exposure that permitted subjects to use this first light as a fixation point. The second point of light appeared for 0·5 sec. and the time interval between the appearance of the first and second lights was varied for the determination of movement thresholds in a manner presently to be related. Since the tachistoscope used was electronically operated, there were no mechanical clicks which subjects might use as additional cues and therefore it was unnecessary to use white noise as in the earlier experiment.

The instructions given to the subjects were designed to indicate the directions in which they should expect the second pinpoint of light to appear in relation to the first. Under one set of conditions the subjects were told that the second spot would appear in any one of *two* possible directions and *only* in these two directions. Thus for example one subject would be told to expect movement only to the North or South (N—S) of the first point of light, another in the E—W direction and so on. The four pairs of directions used were N—S, E—W, NE—SW, and NW—SE, which were allotted randomly to the subjects.

Under the second set of conditions subjects were told that they could expect the position of the second spot of light to be in relation to the first spot of light at any of the *eight* points of the compass: E, NE, N, NW, W, SW, S, SE.

Four threshold determinations were made on each subject, two going from apparent simultaneity to apparent movement and two from apparent movement to apparent simultaneity (i.e. two ascending and two descending trials). The method of limits was employed. For the ascending series, the first presentation was given at a temporal overlap of 245 millisec. between the two stimuli and on successive presentations the overlap was reduced by 35 millisec. until there was no overlap and then the separation was progressively increased by steps of the same size until the subjects gave three successive reports of "movement." The starting point of the determination, an overlap of 245 millisec., produces convincing simultaneity. For the descending series, the first presentation of the two lights was at a temporal separation of 105 millisec. (a duration that for most subjects produces the impression of the apparent movement of one light source) and it was reduced on successive exposures by 35 millisec. until simultaneity (no movement) was finally reported three presentations running.

It would be sensible to consider the ascending series a determination of the threshold of apparent movement, the descending series a determination of the threshold of apparent simultaneity.

The subjects were divided into four subgroups each given somewhat different orders of testing. These are as follows, and subjects were assigned to the subgroups randomly.

| I | II | III | IV |
|---|---|---|---|
| "2" ascending | "2" descending | "8" ascending | "8" descending |
| "8" ascending | "8" descending | "2" ascending | "2" descending |
| "2" descending | "2" ascending | "8" descending | "8" ascending |
| "8" descending | "8" ascending | "2" descending | "2" ascending |

By the use of such a counterbalanced design we are able to determine the effect of the 2- and 8-directional expectancies, the method of computing threshold, the effects of practice or adaptation as one goes from the first to the second half of the experiment, and finally the effect of having a 2-directional or 8-directional expectancy initially.

The subjects were instructed as follows:—

"This is an experiment designed to investigate certain characteristics of apparent visual movement. I want you to look into the eyepiece of this tachistoscope and at intervals I shall show you two points of white light appearing in the black field. Under certain conditions like this (subject shown apparent movement condition until he gets good appearance of apparent movement) you will appear to see not two separate spots, but instead one spot of light appearing to move across the field. Under another set of conditions you will see the simultaneous appearance of two separate and independent spots of white light like this (subject shown simultaneous condition). The second spot of light will appear either vertically upwards or vertically downwards (or whichever condition is being used) and your task is to report to me after each exposure the direction of the second spot and whether it was a case of apparent movement or simultaneity. There may be some instances when you are doubtful; in that case I want you to say so and to report which of the two conditions it most resembles.

Each time before the lights appear I shall say "Ready." At other times you should relax and rest your eyes between exposures. Is that clear? Do you have any questions?"

TABLE I

THRESHOLD OF APPARENT MOVEMENT FOR VARIOUS CONDITIONS EXPRESSED IN MILLISEC. OF TEMPORAL OVERLAP BETWEEN TWO STANDARD LIGHT POINTS

| | *2 directions expected* | *8 directions expected* | *Combined* |
|---|---|---|---|
| All conditions | 130 | 145 | 137 |
| Ascending method | 120 | 113 | 116 |
| Descending method | 140 | 180 | 160 |
| First half determinations | 116 | 144 | 130 |
| Second half determinations | 144 | 146 | 145 |
| 2 directions initially | 137 | 190 | 164 |
| 8 directions initially | 120 | 107 | 113 |
| Ascending: first half | 93 | 102 | 97 |
| Ascending: second half | 145 | 125 | 135 |
| Descending: first half | 140 | 190 | 165 |
| Descending: second half | 140 | 172 | 156 |
| Ascending: 2 initially | 110 | 154 | 132 |
| Ascending: 8 initially | 128 | 72 | 100 |
| Descending: 2 initially | 168 | 223 | 196 |
| Descending: 8 initially | 115 | 138 | 127 |

## RESULTS

Table I presents a summary of the experimental results in terms of the maximum temporal overlap in millisec. between the two points of light which produce the appearance of visual movement. The principal findings may be summarized as follows:—

(1) There is an increase in sensitivity to apparent movement when one operates with an 8-directional expectancy in contrast to a 2-directional one. This difference which is significant at the 5 per cent. level ($t = 2{\cdot}14$, 0·05 level = 2·04) for the thresholds determined by the descending method ceases to be significant when the ascending and descending data are pooled.

(2) There is a marked difference between the results obtained with the ascending and descending methods respectively. The ascending method measures the threshold of apparent movement and the descending method the threshold of apparent simultaneity. When one disregards the expectancy under which subjects worked one finds that the threshold obtained by the ascending method is 116 and by the descending method 160, a difference of 44 which is not significant. If, however, one considers the results from

the first half of the experimental session only, then the difference between ascending and descending thresholds, namely 97 millisec. compared with 165 millisec., a difference of 68 millisec. ($t = 2{\cdot}80$, $0{\cdot}01$ level $= 2{\cdot}62$) becomes significant at the 1 per cent. level. This finding of the effect of halves is not surprising since in our previous study we had found that first half differences tend to wash out in the second half.

(3) When one uses the method of determining the threshold of simultaneity, i.e. descending trials, one obtains a marked difference between 2- and 8-directional expectancies but this difference is all but washed out when one uses the ascending method based on first reports of apparent movement. In view of this it was decided to follow the practice of our previous study and to treat the two sets of data separately and to perform separate analyses of variance on the two sets of data, the ascending and the descending.

(4) Results from the Ascending Method

Of the sources of variance studied the following showed no significant contribution (i) expectancy, i.e. 2 vs. 8 directions, (ii) halves, i.e. first half vs. second half results.

The results did, however, show that with the ascending method the order (i.e. which expectancy was given initially the 2 or the 8) was a significant source of variance. Thus when the 8-condition precedes the 2-condition, we find that the threshold under the 2-condition expectancy is 128 and under the 8-condition expectancy is 72. This difference which is significant at the 5 per cent. level of confidence ($t = 2{\cdot}34$, $0{\cdot}05$ level $= 2{\cdot}13$) is all the more striking since it is in the *opposite* direction to the difference between the 2- and 8-expectancy conditions when all the data are pooled.

(5) Results from the Descending Method

When going from apparent movement to apparent simultaneity the difference between the thresholds under the 2- and 8-expectancy conditions is significant at 5 per cent. level. A markedly greater sensitivity to movement characterizes the 8-directional expectancy. Of the other sources of variance investigated, i.e. order and halves, neither contributed anything significant.

(6) Disregarding the method of determination used, i.e. ascending or descending, the mean threshold obtained under an 8-directional expectancy when it precedes a 2-directional determination is 107 as compared with a threshold of 190 when it follows a 2-directional determination. This difference of 83 millisec. is significant at the 1 per cent. level ($t = 2{\cdot}94$, $0{\cdot}05$ level $= 2{\cdot}00$). More specifically an 8-directional threshold preceding a 2-directional is 107 and following a 2-directional is 190.

(7) One other source of variance which emerged as significant may be summarized as follows: if the ascending determination is done first then the difference between the ascending threshold 97 millisec. and the descending threshold 156 millisec. is 59 millisec., which is significant at the 5 per cent. level ($t = 2{\cdot}59$, $0{\cdot}05$ level $= 2{\cdot}04$); if, however, the descending condition is given initially this difference does not hold up.

## DISCUSSION

In the final paragraph of our earlier paper (Jeeves and Bruner, 1956) we wrote, "We do not, as the reader can see, have a very satisfactory explanation of the findings we have reported. It had been our expectation that quite opposite results would be obtained, and these might well have been handled in conventional informational terms." The challenge of our own dissatisfaction was taken up by Brown (1956) in a note published later the same year. Brown's alternative explanation was couched in "conventional informational terms" and in it he stated that ". . . a reasonable hypothesis concerning the conditions for apparent movement would seem to be: apparent movement tends to be seen if the subject is *able* to discriminate ($d_1$) that the two stimuli are not simultaneous but is *unable* to discriminate ($d_2$) that the two stimuli are not produced by real movement of a single stimulus." This formulation would only be acceptable if, when Brown says "is *unable* to discriminate ($d_2$) that the two stimuli are not produced by real movement of a single stimulus" he means "is unable to discriminate successiveness." This is important because Brown's reference to "real movement" implies a judgement which the subjects in our experiment were at no time required to make since they were told quite explicitly in the experimental instructions that this was an experiment in *apparent movement* and it was made clear that they were reporting the *apparent* movement or *apparent* succession of two points of light. There was no question at any time of real movement of one point of light.

In discussing our earlier explanation Brown drew attention to our second assumption which he stated as "that an expectancy for movement predisposes the subject to see movement" and he criticized it on the grounds that it should operate equally whether going from movement to successiveness or vice versa. This we accept. It is clear that our own earlier explanation itself needs revision even though the predictions made from it for the present experiment have been upheld. We thus find ourselves in the not uncommon situation where an explanation which is unacceptable because of a false assumption nevertheless makes correct predictions for the outcome of an experiment. At the same time Brown's hypothesis which has many attractive features makes predictions contrary to the obtained experimental findings.

In his 1956 explanation Brown drew particular attention to the important factor that "the threshold for a discrimination tends to increase with the information content of any discrimination which has to be made concurrently." In a recent personal communication, Brown has suggested that in the light of the results of the present experiment his hypothesis should be revised so that it now becomes "Apparent movement tends to be seen if the subject is unable to discriminate ($D_1$) that the two stimuli overlap but is able to discriminate ($D_2$) that the second stimulus outlasts the first." According to this reformulation, and assuming that $D_1$ is a difficult discrimination and $D_2$ an easy one, we should predict that with high directional information (the 8-condition) more overlap will be necessary for $D_1$ to occur so that both ascending and descending thresholds will tend to lie at relatively high values of overlap.

Whilst such a prediction does not fit the direction of the results when the 2 first-half-ascending are compared with the 8 first-half-ascending (specifically 110 compared with 72 millisec.) nor when the 2 first-half-descending are compared with the 8 first-half-descending (168 compared with 138 millisec.), this is not serious since these differences fall far short of significance ($t = 1{\cdot}11$ and $t = 0{\cdot}80$, 0·05 level = 2·12). His prediction does, however, fit the direction of the results for the second half (2 ascending 128 millisec., 8 ascending 154 millisec., and 2 descending 115 millisec., 8 descending 223 millisec.) and the descending difference is significant at the 1 per cent. level ($t = 3{\cdot}38$, 0·01 level = 2·92). Moreover the prediction from Brown's new hypothesis fits the overall results which showed a significant difference between 2 and 8 expectancies using the descending method of determination in the direction he would predict.

When the principal findings from the two experiments are considered together we see that with increased directional information (*a*) the movement to successiveness threshold is raised, i.e. movement persists for a greater temporal separation of the two stimuli under the 8 condition and (*b*) the movement to simultaneity threshold is lowered, i.e. movement persists for longer with a greater temporal overlap of the two stimuli.

These results could both be predicted from the following hypothesis: Apparent movement will be seen if the subject is *unable* to discriminate either that the two stimuli overlap or that there is an interval between the two but is able to discriminate that the second stimulus outlasts the first.

The only further assumption that needs to be made is that the latter discrimination is an easy one whereas the two former ones are difficult and it is therefore these which will be most affected by any other discrimination which must be made concurrently (in these experiments whether the second point of light is in one of two or one of eight directions).

The author wishes gratefully to acknowledge the help received from Dr. J. Brown in preparing the present draft of the discussion section of this paper.

## REFERENCES

BROWN, J. (1956). Directional information and apparent movement—A note on the paper by Malcolm A. Jeeves and Jerome S. Bruner. *Quart. J. exp. Psychol.*, **9**, 111–2.

JEEVES, M. A., and BRUNER, J. S. (1956). Directional information and apparent movement. *Quart. J. exp. Psychol.*, **8**, 107–13.

*Manuscript received 24th July*, 1962.

# THE EFFECT OF TOUCH ON A VISUALLY AMBIGUOUS THREE-DIMENSIONAL FIGURE

BY

CHARMIAN SHOPLAND and RICHARD L. GREGORY

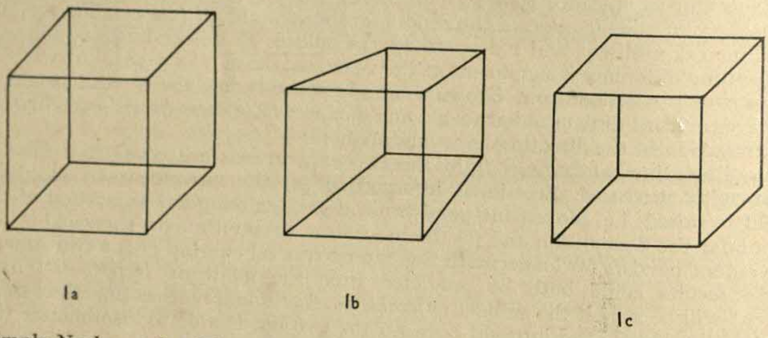
*From the Psychological Laboratory, University of Cambridge*

The effect of haptic (i.e. tactile-kinaesthetic) information on the perception of an ambiguous visual figure was investigated. The figure used was a self-luminous, three-dimensional wire cube viewed in darkness. It was found that visual reversals still took place when the cube was explored with the hands, but the reversal rate was reduced and the time increased during which the cube looked as it "really" was.

## INTRODUCTION

The Necker cube is a well known example of a reversible figure. This simple line drawing (Fig. 1a) is seen to represent a cube in either of two orientations. If, however, perspective or other cues are added (Figs. 1b, 1c), the figure reverses less frequently, and there comes a point when it is no longer ambiguous.

FIGURE 1



A simple Necker cube (Fig. 1a) and two examples (Figs. 1b and 1c) of visual cubes which tend to stabilize its perception.

Mach (1886) pointed out that solid objects could sometimes be seen reversed. Using a three-dimensional line cube we have confirmed this, provided one eye is kept closed and the subject is experienced. However, when we reduce the depth information more drastically, by painting the cube with self-luminous paint, placing it on a black cloth and viewing it monocularly in a completely dark room lined with black paper, we find that even naïve subjects experience reversals spontaneously.

When the cube is seen reversed the face apparently in front looks considerably smaller than the face apparently behind, and none of the corners square. After a little practice we can introduce head movement parallax without the cube returning to normal and this produces a very curious effect. The cube appears to rotate in the opposite direction from the normal one, the apparently nearer parts now moving *with* the head instead of in the opposite direction. The figure undergoes complex (but systematic) shape transformations, giving one the impression that it is made of an elastic material.

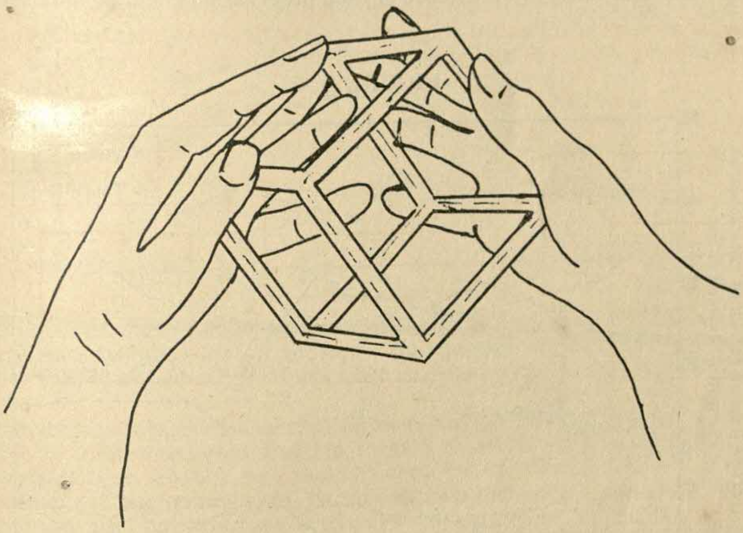
Touching or holding the cube seems to make it more difficult for a naïve subject to see it reversed in the first place, but having once seen it, touching with the fingers or holding in the hands does not prevent reversal.

These preliminary observations led us to carry out the experiment described in this paper, in which we show some effects of touching and holding the cube. One would not expect touch to have any effect on the reversal of a Necker cube drawn on a piece of paper. If touch gave any information, it would be to the effect that the "cube" was in fact flat, but merely looking at a two-dimensional Necker cube one is aware that it is flat—it gives the impression both of flatness and of depth. With the three-dimensional figure it is a different matter—the observer is certain that it is in fact a real cube in front of him and that to touch the front and to touch the back he must put his hands at different distances. One might expect haptic information to stabilize to some extent the perception of this cube, as do visual cues. Alternatively one might suggest that haptic information, being in another modality and not specifically attended to, might be ignored altogether; or one might expect that certain subjects would be greatly affected by haptic information, others hardly at all.

## METHOD

The cube used in this experiment was self-luminous and glowed in the dark. Figure 2 shows its orientation to the (right-eyed) subject holding it, but in the experiment the hands were only faintly visible by the light coming from it. The cube was made of square section brass and had sides 3½ in. long. The subject could cup it conveniently in his hands, simultaneously feeling most of the edges and corners while the cube was fixed in position by a grooved wooden support underneath one edge. He was seated at the table to which the support was fixed, with an eyeshade over one eye and his chin on a rest. The table was covered with black cloth and the walls of the room with black paper.

FIGURE 2



The orientation of the cube to the (right-eyed) subject holding it. The cube was self-luminous in an otherwise dark room. It was wedged into a support which is not shown. The hands were faintly visible by the light of the cube.

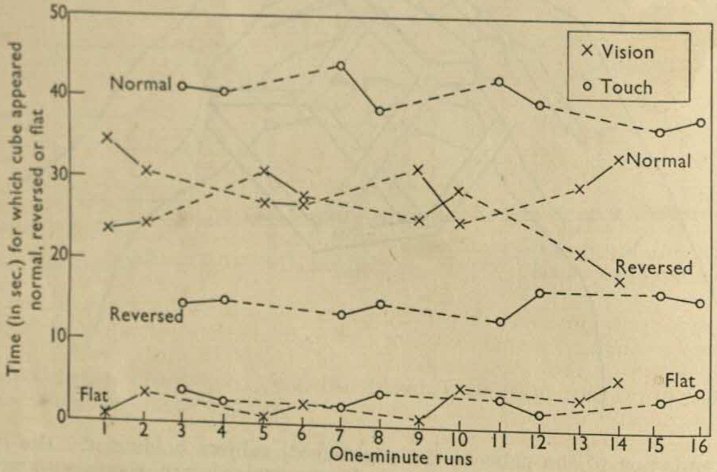
The depth cues which were available to him were as follows: there was a considerable difference in retinal size of the nearer and further faces; he could tell which of two sides crossed in front of the other; the cube was close enough to him for there to be significant changes in accommodation from front to back of it; he could see his own hands faintly by the light coming from the cube. As he used only one eye and his head was kept still by the chin rest, neither binocular disparity nor head movement parallax were available as cues.

Twelve subjects were used in the experiment, all of them being undergraduate or research students in psychology. The subject was first reminded of the Necker figure (all of them had already experienced the reversal of such a figure) and told that he wanted to know whether the solid "figure" in front of him would also reverse. He was asked to note at the beginning of each run whether the cube looked as it "really" was—in which case he was to say "normal," or whether the front and back corners had changed places—in which case he was to say "reversed," or whether the cube looked like a plane figure—in which case he was to say "flat." If anything else was seen he was to say "odd" and explain afterwards.

Each run lasted 1 min. and during that time he was asked to report any changes in the appearance of the cube at the time they occurred. Between each run he spent 10 sec. with his eyes closed. There were eight runs, then a pause for comment and re-activating the cube with a bright light, and then eight more runs. The runs were arranged in double alternations between the condition we call "Vision" and the one we refer to as "Touch." During the "Vision" runs the subject sat with his arms resting on the table and simply looked at the cube; during the "Touch" runs he cupped his hands round the cube and kept both hands moving continuously over its surface. This procedure was explained to him at the beginning of the session and he was told that both active and passive touch were to be used—that is, as well as running his finger tips along the edges of the cube and feeling its lines and corners (trying not to favour any particular lines or corners), he was to feel it with the whole hand.

The reports and comments were recorded on magnetic tape which was played through afterwards so that the times of occurrence of each observation could be obtained.

FIGURE 3



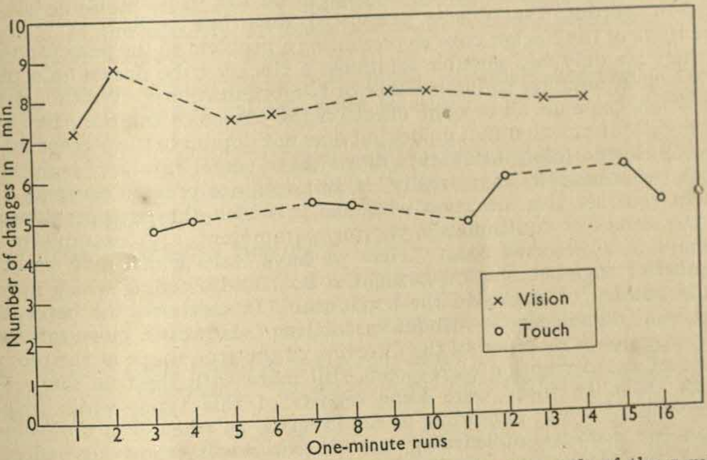
The total time for which the cube appeared "normal," "reversed" or "flat" during each of the 16 1-min. runs of the experiment, averaged for all subjects. When simply looked at, it appeared normal for an average of 30·8 sec. per min., reversed for an average of 25·5 sec. and flat for an average of 3·1 sec. per min. When explored with the hands as well, the corresponding figures are 40·9, 15·4 and 3·6 sec. per min. respectively.

## Results

Figure 3 shows the total time for which the cube appeared "normal," "reversed," or "flat" during each of the 1 min. runs of the experiment, averaged for all subjects. It is clear from inspection alone that when the cube was explored with the hands it looked normal for more of the time and reversed for less of the time than when simply looked at. (Using the Binomial Theorem, $p < 0·004$ in both cases). While exploring with the hands it appeared normal for an average of 40·9 sec. in each 1-min. run and reversed for an average of 15·4 sec. (Binomial Theorem gives $p < 0·004$.) The corresponding figures when simply looking at the cube are 30·8 and 25·5 sec. respectively. (These are not significantly different: Wilcoxon "T" test gives $p > 0·05$). The cube appeared flat for an average of 3·6 sec. on each run when explored with the hands, for 3·1 sec. when simply looked at. This difference is not significant. (Wilcoxon "T" test gives $p > 0{\cdot}05$).

The only "Odd" description of the cube was given by subject 5, on one occasion while using vision alone and on three occasions when using touch also. He reported seeing it in the normal orientation and in depth, but with the kind of distortion that typically occurred when the cube was seen reversed, that is, not as a regular cube but with a marked difference in the sizes of the nearer and further faces. This effect had already been observed by one of the authors (C.D.S.), but had not otherwise been reported by the 20 or so people who took part in the preliminary study. This percept seemed to be independent of any change of attitude on the part of the observer.

FIGURE 4



The number of changes in appearance of the cube during each of the 1-min. runs of the experiment, averaged for all subjects. It changed appearance an average of 7·9 times when simply looked at, 5·3 times when explored with the hands as well.

All the subjects observed that the cube looked distorted this way when reversed. Four subjects commented that fixation seemed to affect reversal (this has previously been observed, (Glen, 1940)); one thought that blinks affected reversal, one that the amount of depth in the figure varied; two thought they could will changes (Glen (1940) has evidence of this also) but said they tried not to do so.

Two subjects thought that touch had no effect at all on what they saw, although they were touching lines they could not see or failing to find with their hands lines they could see. Two subjects found that touching the front corner in particular seemed to make the cube go back to normal if it were reversed, and two found the same with the back corner. (They were both obscuring this corner with the hand, instead of keeping the hand behind the cube as instructed.) One subject found that a corner would sometimes move forward or back with her hand, and another that the part she was touching looked normal while the rest of the cube looked reversed.

Several subjects made the comment that touch gave them a clue to the real orientation of the cube at first, but later they "grew to accept the difference between the visual and the tactile," or found that "touch lost significance." We therefore looked for a decrease in the duration of reports of "normal" from the beginning of the session to the end. Inspection might suggest a very slight trend in the right direction but it is certainly not significant (Friedman Analysis of Variance by Rank gives $p = 0{\cdot}5$). Looking at the results of each subject separately, we found no more indication of variation with time among those who reported it than among the rest.

For all subjects except one there were more changes when using vision only than when using touch as well. (The probability of this distribution is 0·003 on the Binomial Theorem.) The difference in the case of the aberrant subject was nowhere near significant at the 0·05 level (Wilcoxon "T" test), and in other ways his results were typical. Figure 4 shows the number of changes, averaged for all subjects, during each of the 16 1-min. runs of the experiment. Again there is no evidence of any changes with time, nor is there any suggestion of a trend when considering alone the subject who commented that the reversals became more frequent with time.

The reversal rate in our experiment is less in both conditions than is generally reported for two-dimensional Necker cubes, our average rates being 7·9 per min. with vision only and 5·3 with touch also, compared with average rates ranging from 15 to 30 in the literature (e.g. Tussing, 1941; Glen, 1940).

## DISCUSSION

This experiment shows that haptic information does not prevent reversals of an ambiguous figure, nor does it prevent the figure from looking like a "meaningless geometrical pattern." It could well have been the case that when, say, the front of the cube was located by touch it would always appear to be the front visually, but in fact the cube seemed to split into two objects, one visual, the other tactile.

We may think of the Necker cube as presenting a problem to the perceptual mechanism for which there are only two possible solutions: a Necker cube drawn on a plane surface is usually bi-stable, seeming to lie in either of two orientations. We thought it possible that touch would have an all-or-none effect on the choice of solution, but we find that touch can provide information that guides but does not dominate the perceptual mechanism in its choice. Haptic information cuts down the reversal rate and increases the time during which the cube looks as it "really" is, but does not prevent reversal.

We might contrast this situation with one in which the perceptual problem is to decide on the values of continuously varying parameters. An example of this is the apparent shape of a distorted room. Here we have walls which may be at any of an infinite number of angles to the line of sight, a floor and a ceiling which may be at any of an infinite number of angles to the horizontal. In this case the perceptual system makes a decision and sticks to it without vascillating. However, given information from touch, the perception is modified in the direction of the true shape of the room, and with more experience the perception corresponds still more with the true shape (Kilpatrick, 1954). With the cube there were some reports of this type—while touching it, it appeared reversed, but not extending as far in space as when seen unreversed. There was however, no evidence of learning, nor do our results show any effect of haptic experience of the cube on its perception when using vision alone. From the reports of the subjects (but not from the numerical results) we might suggest that they learned not to use but to ignore haptic information as time went on. Perhaps if their task had been to ascertain which way round the cube in fact was, touch would have had more effect.

There were also reports that touching the front or back corner seemed to cause an immediate return to the normal orientation. Witkin *et al.* made a similar finding when putting vision and hearing into conflict: when attempting to judge the direction of a speaker's voice, opening the eyes could make the voice appear to change direction by about 30 degrees instantaneously (Witkin, Wapner and Leventhal, 1952). However, in their experiment, the auditory stimulus was unambiguous and a more comparable situation is that of Warren and Gregory (1958), where a word repeated on a loop of tape was found to change spontaneously to other words or sounds.

One of us (C.D.S.) was supported by a grant from the Medical Research Council.

## REFERENCES

GLEN, J. S. (1940). Ocular movements in reversibility of perspective. *J. gen. Psychol.*, **23**, 243–81.

KILPATRICK, F. P. (1954). Two processes in perceptual learning. *J. exp. Psychol.*, **47**, 362–70.

MACH, E. (1886). *The Analysis of Sensations.* (Translation of 5th edition), New York, 1959.

SIEGAL, S. (1956). *Nonparametric Statistics for the Behavioral Sciences.* New York.

TUSSING, L. (1941). Perceptual fluctuations of illusions as a possible fatigue index. *J. exp. Psychol.*, **29**, 85–8.

WARREN, R. M., and GREGORY, R. L. (1958). An auditory analogue of the visual reversible figure. *Amer. J. Psychol.*, **71**, 612–3.

WITKIN, H. A., WAPNER, S., and LEVENTHAL, T. (1952). Sound localisation with conflicting visual and auditory cues. *J. exp. Psychol.*, **43**, 58–67.

*Manuscript received 4th September, 1962.*

# PROCEEDINGS OF THE EXPERIMENTAL PSYCHOLOGY SOCIETY, 1963

*3rd–5th April*, 1963. Extended Meeting at Cambridge.

*1st Session:* "On the representation of choice reactions by a random walk," by D. R. J. Laming.* "The relationship of reaction time and alpha blocking in normal and severely subnormal subjects," by B. Hermelin and P. H. Venables. "The origin of the after-effect of movement," by M. Pickersgill.* "Repulsion and adaptation in the figural after-effect," by J. Wilson.*

*2nd Session:* Symposium on studies of cognitive learning, with contributions from M. S. Halliday,* M. H. Sheldon,* A. W. Still,* and A. J. Watson. "Acoustic confusions in memory," by R. Conrad.

*3rd Session:* "Ganglion cells in the rabbit's retina signalling direction of motion," by H. B. Barlow. "Studies in the two-point threshold," by O. L. Zangwill and B. R. Gomulicki.*

*4th Session:* Symposium on reading aids for the blind, with contributions from M. B. Clowes* and P. W. Nye,* introduced by J. A. Leonard. "Pattern perception in the stabilized retinal image," by C. R. Evans.*

*9th–11th July*, 1963. Extended Meeting at Oxford.

*1st Session:* "A comparison of recall and recognition," by J. Brown. "Apparent movement in repetitively patterned fields," by M. E. Wilson.* "Feedback control of limb-movement and Fitt's Law," by E. R. F. W. Crossman and E. J. Goodeve.*

*2nd Session:* "Multi-channel listening in the four-eared man," by N. Moray. "Auditory unmasking," by M. Treisman. "Problems in the measurement of perceived temporal order," by I. J. Hirsh.*

*3rd Session:* Two papers on stimulus analysing mechanisms, by N. J. Mackintosh* and N. S. Sutherland. "Experimental studies of 'sharing' behaviour," by J. T. Lanzetta.*

*4th Session:* "Electrical stimulation of the brain in Barbary Doves," by D. M. Vowles and D. Harwood.* "A possible mechanism underlying phototaxis in amphibia," by W. R. A. Muntz. "Plasticity of phototactic behaviour in frogs and terrapins," by N. Mrosovsky. "Perceptual skills as components of intelligent behaviour," by A. Elithorn and Myfanwy Kerr.* "Perceptual and intellectual skills," by D. N. Lee.*

*6th–8th January*, 1964. 16th Annual General Meeting at Birkbeck College, London.

*1st Session:* "Neurochemical regulation of learning," by R. B. Ross.* "The relation between stimulus intensity and operant response rate," by J. A. Gray.* "Is seminal ejaculation exhausting or satiating or neither?" by L. J. Herberg.

*2nd Session:* "A theory concerning nervous changes during the preparation for hibernation based on behavioural tests of motivation of dormice and other hibernators and its possible relevance to artificial hypothermia in rats," by N. Mrosovsky. "An application of the method of paired comparisons to the study of the Delboeuf illusion," by J. A. Keats.* "Evidence for a physiological conception of human language," by Davis Howes.

*3rd Session:* "Experiments on the syntactic aspects of negation," by G. A. Miller. "Experiments on the pragmatic aspects of negation," by P. C. Wason. "Time discrimination and the indifference interval," by M. Treisman.

*4th Session:* "Capacity limitations and the span of attention," by D. Jones.* "Selection and order in immediate memory," by G. J. Harrison.*

*5th Session:* "Comparison of post-rotatory sensation and nystagmus," by A. J. Benson. "An 'adaptive threshold' hypothesis to account for individual differences in the duration of visual and labyrinthine after-sensations," by J. T. Reason.*

* Speakers by invitation.

*Committee*, 1964: President: Professor O. L. Zangwill.
Editor: Professor O. L. Zangwill.
Ordinary Members of the Committee: Dr. R. J. Audley (Hon. Secretary), Dr. R. Davis (Hon. Treasurer), Dr. J. Gedye, Mr. R. L. Gregory, Dr. C. I. Howarth, Dr. N. Moray, Dr. N. O'Connor, Dr. P. Venables (Asst. Hon. Secretary), Dr. L. Weiskrantz.

The following have been elected into membership of the Society:

A. D. Baddeley, Ph.D., M.R.C. Applied Psychology Research Unit, 15 Chaucer Road, Cambridge.
G. S. Brindley, M.D., Trinity College, Cambridge.
M. B. Clowes, Ph.D., M.R.C. Psycholinguistics Research Unit, Institute of Experimental Psychology, Oxford.
D. W. J. Corcoran, Ph.D., M.R.C. Applied Psychology Research Unit, 15 Chaucer Road, Cambridge.
M. Hammerton, B.Sc., M.R.C. Applied Psychology Research Unit, 15 Chaucer Road, Cambridge.
I. M. Hughes, Ph.D., University of Manchester, Manchester, 13.
J. G. Ingham, Ph.D., M.R.C., Social Psychiatry Research Unit, Institute of Psychiatry, Maudsley Hospital, Denmark Hill, London, S.E.5.
Myfanwy O. Kerr, B.A., Royal Free Hospital, Gray's Inn Road, London, W.C.1.
N. J. S. M. Mackintosh, B.A., Institute of Experimental Psychology, 1 South Parks Road, Oxford.
Irene Martin, Ph.D., Institute of Psychiatry, Maudsley Hospital, Denmark Hill, London, S.E.5.
Mary J. Pickersgill, Ph.D., Department of Psychology, University of Leicester, Leicester.
R. G. A. Stretch, Ph.D., Department of Psychology, Queen's University, Belfast.
J. A. Wilson, B.Sc., Department of Psychology, University of Bristol, Bristol 8.

*Visiting Foreign Members*

Davis Howes, Ph.D., Psychological Laboratory, Downing Street, Cambridge.
Professor G. A. Miller, Institute of Experimental Psychology, 1 South Parks Road, Oxford.
Professor L. Postman, Department of Psychology, University College London, Gower Street, London, W.C.1.

# PUBLICATIONS RECEIVED

## JOURNALS

*Psychologica*. Annual Periodical published by the Department of Psychology and Institute of Psychology, Faculty of Arts, Comenius University, Czechoslovakia. 1961. English summaries. Pp. 181.

*Acta Biologiae Experimentalis*. Vol. XXIII, No. 2 and No. 3. 1963. Editor-in-chief: Jerzy Konorski. Managing Editor: Stefan Brutowski, The Nencki Institute of Experimental Biology, Warsaw. Published quarterly in issues of 80 pages approx. Most papers in English, others in French or German. Subscriptions to Export and Import Enterprise "Ruch" Warsaw, Wilcza 46.

*Neuropsychologia*. Vol. I, Nos. 1 and 2. Edited by D. Denny-Brown, Cl. Faust, H. Hécaen, A. R. Luria, K. H. Pribram, H. L. Teuber and O. L. Zangwill. Published quarterly. Oxford: Pergamon Press. Annual Subscription Rates: For libraries etc. £10 ($30.00); for private individuals £5 ($15.00). Papers in English, French or German.

*Studia Phonologica*. Vol. II, 1962. Chief Editor: Hisanosuke Izui. Institution for Phonetic Sciences, University of Kyoto, Japan. Papers in English or with English summaries.

*Kybernetik: A Journal dealing with the Transmission and Processing of Information as well as with Control Processes in Both Organisms and Automata*. Vol. I, Part 3, 1961. Edited by H. B. Barlow, M. Halle, B. Hassenstein, W. D. Keidel, I. Kohler, K. Küpfmüller, H. Mittelstaedt, W. Reichardt, W. A. Rosenblith, J. F. Schouten, M. Schützenberger, K. Steinbuck and N. Weiner. Published by Springer-Verlag, Heidelberger Platz 3, Berlin-West. 6 numbers per volume. Approx. DM80 or $20.

## MONOGRAPHS

*Odor and Olfaction: A Bibliography, 1948–1960*. By K. M. Michels, D. S. Phillips, R. H. Wright and J. Pustek, Jr. Perceptual and Motor Skills Monograph Supplement 6-V15. 1962. Pp. 55. $3.00.

*Correlational Methods in Research on Human Learning*. By Winton H. Manning and Philip H. Du Bois. Percept. Motor Skills Monogr. Supp. 3-V15. 1962. Pp. 35. $2.00.

*Passages from the "Idea Books" of Clark L. Hull*. Percept. Motor Skills. Monogr. Supp. 9-V15. 1962. Pp. 81. $3.00.

*Relation between Effort, Tension Level, Skill and Performance Efficiency in a Perceptual Motor Task*. By Robert G. Eason. Percept. Motor Skills Monogr. Supp. 2-V16. 1963. Pp. 20. $1.00.

*Eye Fixation and Recall of Pictures as a Function of GSR Responsivity*. By Lester Luborsky, Barton Blinder and Norman Mackworth. Percept. Motor Skills Monogr. Supp. 5-V16. 1963. $1.00.

*An Interpretation of the Phenomenological Structure of Perceived Motions*. By Thomas B. Mulholland. Percept. Motor Skills Monogr. Supp. 2-V17. 1963. Pp. 18. $1.00.

*Aversive Conditioning Methodology in Animal Research*. By Melvin L. Goldstein. Psychological Reports Monograph Supplement 5-VII. 1962. Pp. 28. $1.00.

*Continuous Recording of Spontaneous Runway Behavior: A Normative and Methodological Study*. By A. C. Pereboom. Psychol. Rep. Monogr. Supp. 2-VII. 1962. Pp. 15. $1.00.

*Parsimony in Psychology*. By William F. Battig. Psychol. Rep. Monogr. Supp. 3-VII. 1962. Pp. 16. $1.00.

*The Quick Test (QT): Provisional Manual*. By R. B. Ammons and C. H. Ammons. Psychol. Rep. Monogr. Supp. 1-VII. 1962. Pp. 53. $2.50.

*Psychology of the Scientist: V. Three Experiments in Experimenter Bias*. By Robert Rosenthal and Kermit L. Fode. Psychol. Rep. Monogr. Supp. 1963. Pp. 20. $1.00.

*Constant vs Varied Serial Order in Paired-Associate Learning*. By William F. Battig, Sam C. Brown and Douglas Nelson. Psychol. Rep. Monogr. Supp. 6-VI2. 1963. $1.00.

*Personal Structures: I. A Basis for Personal Structures*. By Maynard W. Shelly. Psychol. Rep. Monogr. Supp. 1-VI3. 1963. Pp. 17. $1.00.

*Untersuchungen über den Grad der subjectiv gegebenen Beanspruchung bei körperliche Arbeit.* By Hugo Schmale, Heinz Schmidtke, and Adolf Vukovich. Cologne and Opladen: Westdeutscher Verlag. Forschungsberichte des Landes Nordrhein-Westfalen, No. 1261. 1963. Pp. 63. DM27.80.

## BOOKS

Review in a later issue is not precluded by notice here

*Psychology: A Study of a Science. Vol. 5. The Process Areas, the Person, and Some Applied Fields: Their Place in Psychology and in Science.* Edited by S. Koch. London: McGraw Hill. 1963. Pp. xii + 967. 104s. 6d. *Vol. 6. Investigations of Man as Socius: Their Place in Psychology and the Social Sciences.* Edited by S. Koch. London: McGraw Hill. 1963. Pp. xii + 791. 97s.

*Mental Health Book Review Index.* Vol. 8, 1963. Whole No. 13. With an Editorial: Changes in the concept of "scientific literature." Pp. xv + 68. Sponsored by the World Federation for Mental Health, the Research Center for Mental Health, New York University, and others. Annual subscription $5.00. The Index, c/o Miss Lois Afflerbach, Paul Klapper Library, Queens College, Flushing 67, N.Y.

*Passing Examinations.* By Clifford Allen. London: Macmillan. 1963. Pp. xiii + 146. 12s. 6d.

*Birth to Maturity: A Study in Psychological Development.* By Jerome Kagan and Howard A. Moss. London: Wiley. 1963. Pp. xiii + 364. 65s.

*Personality: Dynamics and Development.* By Irving Sarnoff. London: Wiley. 1963. Pp. xii + 572. 59s.

*Story Sequence Analysis.* By Magda B. Arnold. London: Columbia University Press. 1963. Pp. vii + 287. 56s.

*The Narrow Pass: A Study of Kierkegaard's Concept of Man.* By George Price. London: Hutchinson. 1963. Pp. 224. 30s.

*Psychology in Belgium.* By Joseph Nuttin. Louvain: Nauwelaerts. 1961. Pp. 77. FB 70- ($1.60).

*Les Reactions Psycho-Physiolgiques à la Douleur.* By A. E. McKenna. Louvain: Nauwelaerts. 1963. Pp. 126. FB 120.

*Motivation as Related to Personality.* By Dorothy Rethlingshafer. London: McGraw Hill. 1963. Pp. ix + 388. 62s.

*Problems of Psychiatry and Neurology.* Edited by I. F. Sluckevski. Oxford: Pergamon Press. 1963. Pp. viii + 376.

*Psychology and Religion: An Introduction to Contemporary Views.* By G. Stephens Spinks. London: Methuen. 1963. Pp. xv + 221. 28s.

*Committees: How They Work and How to Work Them.* By Edgar Anstey. London: Allen and Unwin. 1962. Pp. 116. 15s.

*Conceptele Figurale.* By E. Fischbein. Bucharest: Editura Academei Republicii Populare Romine. 1963. Pp. 473. Lei 16.30.

*Computer Simulation of Personality: Frontier of Psychological Theory.* Edited by Silvan S. Tomkins and Samuel Messick. London: Wiley. 1963. Pp. x + 325. 40s.

*The Quest for the Father.* By Phyllis Greenacre. London: Bailey & Swinfen. 1963. Pp. 128. 27s.

*Social Foundations of Industrial Psychology.* By W. T. Adiseshiah. India: United Printers, Madras. 1963. Pp. viii + 172. 12s. 6d.

*Philosophy of Science: The Delaware Seminar. Volume I 1961–62.* Edited by Bernard Baumrin. London and New York: Interscience Publishers (Wiley). 1963. Pp. xvi + 370. 75s.

*An Introduction to Psychotherapy.* By Sidney Tarachow. London: Bailey & Swinfen. (New York: International Universities Press). 1963. Pp. viii + 376. 68s.

*Experiments with Drugs.* Edited by H. J. Eysenck. Oxford: Pergamon. 1963. Pp. xii + 421. 100s.

*The Prediction and Control of Behavior: A Shorter Introduction to Psychology.* By J. D. Keehn. London: Constable. 1963. Pp. 268. 45s.

*Creativity and Psychological Health.* By Frank Barron. London: Van Nostrand. 1963. Pp. xi + 292. 51s.

*Psychology: The Study of Behavior.* By Paul Swartz. London: Van Nostrand. 1963. Pp. xix + 451. 54s. 6d.

*Soundmaking: The Acoustic Communication of Emotion.* By Peter F. Ostwald. Springfield, Illinois: Thomas. 1963. Pp. xii + 186. $7.00.

*The Physiology of Insect Senses.* By V. G. Dethier. London: Methuen. 1963. Pp. ix + 266. 42s.

*Scientific Principles of Psychology.* By Donald J. Lewis. New Jersey and London: Prentice-Hall. 1963. Pp. xi + 580. 64s.

*Digital Computers in Research: An Introduction for Behavioral and Social Scientists.* By B. F. Green. London and New York: McGraw Hill. 1963. Pp. xii + 333. 83s. 6d.

*Alfred Adler: The Man and His Work.* By Bertha Orgler. Third Edition (revised and enlarged). London: Sidgwick and Jackson. 1963. Pp. xix + 266. 30s.

*Readings in Mathematical Psychology.* Vol. I. Edited by R. Duncan Luce, Robert R. Bush and Eugene Galanter. New York and London: Wiley. 1963. Pp. ix + 535. 68s.

*Clinical Psychology: An Introduction to Research and Practice.* By N. D. Sundberg and L. E. Tyler. London: Methuen. 1963. Pp. xviii + 564. 50s.

*Critical Essays on Psychoanalysis.* Edited by Stanley Rachman. Oxford: Pergamon Press. 1963. Pp. xiii + 284. 45s.

*Color: A Guide to Basic Facts and Concepts.* By R. W. Burnham, R. M. Hanes and C. James Bartleson. London and New York: Wiley. 1963. Pp. xi + 249. 70s.

*Psychopharmacological Methods: Proceedings of a Symposium on the Effects of Psychotropic Drugs on Higher Nervous Activity, Prague, 1961.* Edited by Z. Votava, M. Horváth and O. Vinař. Oxford: Pergamon. 1963. Pp. 360. £5.

*Scientific Principles of Psychology.* By Donald J. Lewis. London and New Jersey: Prentice-Hall. 1963. Pp. xi + 580. 56s.

*Bargaining Behavior.* By Lawrence E. Fouraker and Sidney Siegel. New York and London: McGraw Hill. 1963. Pp. ix + 308. 69s. 6d.

*Syndromes of Psychosis.* By Maurice Lorr, C. James Klett and Douglas M. McNair. Oxford: Pergamon Press. 1963. Pp. x + 286. 70s.

*The Physiological Basis of Mental Activity: Proceedings of a Symposium held in Mexico City, 1961.* Edited by Raúl Hernández Peon. Supplement No. 24 to *Electroenceph. clin. Neurophysiol.* London: Elsevier. 1963. Pp. ix + 283. 182 illus. 90s.

*The Design of Electric Circuits in the Behavioral Sciences.* By Tom N. Cornsweet. London and New York: Wiley. 1963. Pp. xi + 329. 68s.

*Mankind Behaving: Human Needs and Material Culture.* By James K. Feibleman. Springfield, Illinois: Thomas. 1963. Pp. vii + 275. $10.50.

*Handbook of Counseling Techniques.* Edited by Ernest Harms and Paul Schreiber. Oxford: Pergamon Press. 1963. Pp. vii + 506. 84s.

*Statistical Concepts: A Program for Self-Instruction.* By Celeste McCollough and Loch van Atta. London: McGraw Hill. 1963. Pp. xv + 367. 46s. (Paperback Edition 31s.)

*Progress in Brain Research. Vol. I. Brain Mechanisms.* Edited by G. Moruzzi, A. Fezzard and H. H. Jasper. London: Elsevier. 1963. Pp. xix + 493. 140s.

## PAPERBACKS

*Contemporary Research in Learning.* Edited by John R. Braun. Insight Book No. 10. London: Van Nostrand. 1963. Pp. vii + 229. 15s. 6d.

*Primate Social Behavior.* Edited by Charles H. Southwick. Insight Book No. 14E. London: Van Nostrand. 1963. Pp. viii + 191. 15s. 6d.

*Research in Psychopathology.* Edited by Herbert C. Quay. Insight Book No. 11. London: Van Nostrand. 1963. Pp. viii + 216. 15s. 6d.

*Transfer of Learning.* Edited by Robert F. Grose and Robert C. Birney. Insight Book No. 9E. London: Van Nostrand. 1963. Pp. vi + 194. 14s.

*Psychology in the Wry.* Edited by Robert A. Baker. Insight Book No. 12. London: Van Nostrand. 1963. Pp. xii + 170. 14s.

*Experimental Psychical Research.* By Robert H. Thouless. London: Penguin Books. 1963. Pp. 148. 3s. 6d.

*The Science of Animal Behaviour.* By P. L. Broadhurst. London: Penguin Books. 1963. Pp. 143. 3s. 6d.

*Clinical Psychiatry for the Layman.* By F. J. Fish. Bristol: John Wright. 1963. Pp. 68. 8s. 6d.

# BOOK REVIEWS

*Annual Review of Psychology*, Vol. 14. Edited by P. R. Farnsworth. Palo Alto, California: Annual Reviews Inc. 1963. Pp. x + 454. $9.00.

The basic idea of the *Annual Review* is such a good one that it is difficult to imagine what a really bad edition would be like. Even when the reviews provide no more than an annotated bibliography of the year's work, sometimes dully written and poorly selected, they are still extremely useful as guides to the new literature, more convenient, although less exhaustive, than Psychological Abstracts. However, the Review offers an opportunity to do much more than this. It presents an opportunity for commenting on, as well as reporting, the work done. Indeed the main criticism of the *Annual Review of Psychology* is that some of the contributors report in equal detail the trivial and the important, failing altogether to comment on the quality of the work. By far the most satisfactory reviews are produced by those authors who evaluate the progress of research over the whole of their chosen field.

By these criteria the most satisfactory review this year is the one by Chapanis on Engineering Psychology. He is properly and discriminatingly appreciative of a few good studies such as Geldard's work on communication via the skin. He devotes more space than other reviewers to methodological and theoretical papers. But his most important contribution is that he has taken the first practical step towards reducing the avalanche of third class papers which is making so many of our journals unreadable. He has devoted two pages of his review to castigating two impossibly bad papers and one good, but flawed, paper. At present the penalties for publishing poor work are very slight, but if the *Annual Review* were to encourage other reviewers to follow Chapanis' example it would powerfully discourage careless overpublication.

The more one thinks about this idea the more exciting it becomes. Reviewers of future *Annual Reviews* may have to consider whether the worst type of publications have been sufficiently discouraged. Chapanis is mainly concerned about the publication of meaningless or unreliable data. But would it not be just as useful to discourage the piecemeal publication of a series of experiments which should be gathered together in a single paper, or the development of private languages, often mathematical or symbolic, to express ideas which are more usefully and simply expressed in words? If the contributors to the *Annual Review* are to act as Lord High Executioners, in this way, perhaps the editors should think seriously about compiling a "little list" of the kind of crime they think should merit this type of punishment.

None of the other reviews contains such inflammatory material. They are of the usual mixed quality. As usual no one has been able to report a really major discovery, although this does not necessarily mean that none have been made. At the other end of the scale too many of the reviewers report trivial work so uncritically as to be misleading. For example, the chapter on Audition reports a study using the Bekesy audiometer in which higher thresholds were obtained when the subject was instructed "never to let the tone fade away completely" compared with the threshold obtained in the normal way with the stimulus fluctuating between audibility and inaudibility. Such trivial results should either be ignored or criticized, but surely not reported without comment. The same chapter reports, in no greater detail, the infinitely more important discovery that the Bekesy model of the cochlea shows displacement patterns which correspond to the "missing fundamental." Up to now the evidence has seemed to indicate that the audibility of the "missing fundamental" must be due to neural mechanisms. Bekesy's new discovery means that this evidence will now have to be re-examined.

In the chapter on Personality, barely significant correlations of about 0·15–0·3 are reported with depressing frequency. Some of these have apparently been extracted from a matrix of correlations, most of which are not significant. Correlations of this order are hardly worth reporting unless supported by some attempt to discover why the correlation exists or at least to discover a way of increasing it.

One of the better chapters is that by E. Gibson on perceptual learning. This is not surprising since she is responsible for a high proportion of the studies in this area and is one of the inventors of the "visual cliff" apparatus which is, perhaps, the simplest and most flexible method of investigating the effect of experience on behaviour. Surprisingly, there is no mention in this chapter of the work of Lawrence or Sutherland, who have been thinking about, and working on, problems of perceptual learning from a theoretical position rather different from Gibson's.

In physiological psychology two new preparations, the split brain preparation of Sperry and gastric injection preparation of Epstein, seem likely to yield much valuable data. But perhaps the most satisfying piece of work in this field is the purely physiological work of Donner and Rushton on the colour responses of "excitation pools" in the frog retina. At last physiological techniques are being developed which may be more useful than the purely behavioral techniques which, over the past two hundred years or so, have been so much more significant than any physiological work on this subject. Incidentally the chapter on Colour Vision does less than justice to the elegant theoretical work of Brindley in which he deduces, from purely behavioural data, that only three pigments are involved in colour vision, and that rhodopsin is not one of them.

In the past the *Annual Review* has been an indispensible addition to any psychological library. In the future, if Chapanis's lead is followed, it could perform an infinitely more useful function. It could become the guardian of our scientific integrity.

C. I. Howarth.

*Interhemispheric Relations and Cerebral Dominance*. Edited by Vernon B. Mountcastle. London: Oxford University Press (The Johns Hopkins Press). Pp. x + 294. 1962. 60s.

This important symposium was held at the Johns Hopkins University Medical School in April, 1961, and should be of great value to psychologists interested in brain mechanisms and behaviour. Inevitably, a good deal of interest centres on the "split-brain" preparation and its value in neurological analysis, though it is noteworthy that Dr. Sperry does not now regard the "split-brain" animal as so completely "split" as we have hitherto been led to suppose. In addition to Dr. Sperry, Dr. R. E. Myers and Dr. J. L. de C. Downer have important papers on this topic. Also of interest is Dr. Ettlinger's comparison of the effects of successive unilateral ablations of the posterior parietal cortex with those produced by a single bilateral removal. The results lead him to conclude that in the monkey, there is a predominantly crossed relation between each posterior parietal area and performance of the contralateral hand in tactile shape discrimination. Dr. Mishkin contributes a thoughtful (and thought-provoking) comment on the possible link between inter-hemispheric integration in the monkey and cerebral dominance in man.

The papers on human neurological issues are more varied, but concentrate mainly on the differences which appear to exist at the psychological level between the effects of broadly equivalent lesions of the two hemispheres. Dr. Teuber and Dr. Weinstein review some of the Bellevue work from this standpoint, though do not add a great deal to the material already published. Dr. Brenda Milner describes some interesting laterality effects in audition observed in cases of temporal lobectomy which may prove of some importance. Of particular interest is her adaptation of a dichotic listening technique originally devised by Broadbent for quite other purposes. Dr. Hécaen presents a substantial review of the differences in clinical symptomatology which appear to be associated with the laterality of the lesion, placing particular emphasis on right-sided lesions in relation to disorders of spatial orientation and certain aspects of visual skill. His interpretation is however sharply challenged by Dr. Denny-Brown in a comment of characteristic vigour. Other contributors to this part of the symposium include Dr. H. G. Wolff, Dr. M. Critchley and Dr. Arthur Benton.

The work assembled in this volume makes it clear that some revision of the concept of cerebral dominance, which goes back at least to Hughlings Jackson, is overdue. At the same time, it is unfortunate that we are given no very clear line as to how this revision should proceed or what types of inquiry are likely to prove most fruitful. In spite of this most welcome symposium, cerebral dominance remains an enigma.

O. L. Zangwill.

*Thought and Language*. By S. Vygotsky. Edited and translated by Eugenia Hanfmann and Gertrude Vakar. New York and London: M.I.T. Press and John Wiley & Sons Inc. 1962. Pp. xxi + 168. With separate comments by Jean Piaget. 38s.

This is a book of quite unusual interest and quality. No reader can fail to be constantly struck by the originality and penetration of Vygotsky's observations and analysis. Of the circumstances of his life and work it need only be said that his age in 1917 was 21, that he was deeply inspired by the liberal educational doctrines of Tolstoy, that he worked in Moscow throughout the 1920's and that this book was published soon after his death in 1934. The original text is said to have been repetitious, polemical and involved and the present translators decided to condense considerably in the interests of coherence. So far as one may judge, this has been very well done and much gratitude is owing to the initiative of Professor Luria and to the translators for making the work available for the first time at first hand to Western readers.

Vygotsky's aim is the analysis of the relations of thought and language—of concepts and words—especially in their genetic aspect. The empirical work done by him and his colleagues on which the analysis is based is, unfortunately, only very incompletely sketched in and according to Professor Luria no complete report was ever published. The discussion largely centres round two topics: first, the character and function of inner speech and its development in the child *via* egocentric speech, and secondly the complicated stages by which the capacity to form and use true concepts is attained. Vygotsky starts by criticizing Piaget's early views as to the significance of egocentric speech. (With the book is included a separate pamphlet by Professor Piaget in which he comments on these criticisms, maintaining that to the extent they are justified they largely apply only to the views expressed in his first two books, *The Language and Thought of the Child* and *Judgement and Reasoning in the Child*). Piaget, Vygotsky holds, is at fault in treating egocentric utterance as a primary and merely autistic manifestation which atrophies with the later development of socialized speech. On the contrary, speech is social—if in a primitive sense—from the very outset, and only as its communicative functions begin to be elaborated does its egocentric use come fully into being. Its fate is not simply to fade away but, passing through a series of stages in which its syntactical and semantic structure progressively changes, it becomes internalized as inner speech. As the child's overt egocentric remarks grow fewer, they become increasingly condensed and inscrutable to any other person unacquainted with their internal context. Just as in the exchange of thought between two well-acquainted adults only key elements of verbal currency need be used, so inner language is the vehicle of thought, though it may incorporate little in the way of explicit verbal imagery.

Thought, however, is not identical with language and the two have separate origins in the early development of the child. Both phylogenetically and ontogenetically, on the one hand there is a pre-linguistic logic of action: on the other, speech develops before the time when it can have any truly conceptual background. Vygotsky's account of the evolution of true concepts in the individual is based upon extensive experiments by himself and Sahkorov using materials which are familiar in the Vygotsky Sorting Test. It transpires that true concepts appear only after a series of stages in which classification of objects takes place in accordance with more primitive principles. First comes the formation of "*congeries*" or "heaps." These are built on a vague syncretic basis or as a result of some feature in the immediate perceptual field. The next stage is the occurrence of *complexes*. Here the principle of association is at a superficial factual level and may vary as the classification proceeds. In its most highly developed form complex-formation may lead to a grouping outwardly indistinguishable from one which might be made by an adult acting in accordance with a true concept. Closer enquiry reveals, however, that no true conceptual principle is at work. The final stage in the development of conceptual thinking, which occurs only in early adolescence, requires not only the abstraction of common properties but also the application of the abstracted elements to fresh concrete situations or possibilities. This development can only take place *pari passu* with the evolution of language to the point where appropriate verbal meanings become available.

No brief summary can do justice to the remarkable freshness and stimulating quality to be found on almost every page. The extended analysis of complex-formation, as exemplified in the classificatory activity of the child, is made to illuminate a number of apparently anomalous and irrational features of adult thinking. Those, again, who are interested in the breakdown of intellectual function in cerebral pathology or in psychotic illness will find this part of Vygotsky's work suggestive and helpful. This is not a work to which one may look for a coherent body of final conclusions based on explicit empirical data. It is rather a collection of promising starting points as, in fact, it has been for contemporary Russian workers in this field. It should be not merely read but carefully studied by everybody interested in child development or in the psychology of language and thinking.

R. C. OLDFIELD.

*Porpoises and Sonar.* By W. N. Kellogg. London: Phoenix Science Series, University of Chicago Press. 1963. Pp. xiv + 177. 10s. 6d.

Everyone is now familiar with the work of Griffin on the mechanism whereby bats not only fly dexterously in the dark amongst the interlacing branches of trees or similar obstacles, but actually locate and catch their aerial prey solely by a system of sonar. In the bat this consists of a series of ultrasonic pulses which can be beamed and the returning echo received with such precision as to render sight superfluous. Not only are bats able to do this; they are also able to distinguish between echoes returning from their own pulses and those from other sources; and they are also extremely adept at discriminating such echoes from louder noises. It now appears that porpoises and probably many other Cetacea also orient in this manner. Dr. Kellogg has put his work on this subject into an attractive and readable book entitled *Porpoises and Sonar.* Dolphins and porpoises being smaller and somewhat less overpowering than whales are among the most attractive, graceful and intelligent of marine animals. The book outlines many matters of behavioural and biological interest in these astonishing creatures. It includes an assessment of those rare, but seemingly authentic, instances in which dolphins will play with swimming children and with divers; sometimes even swimming beneath their legs so as to carry them a short distance before dropping them again. But the greater part of the book is devoted to the clicking and whining sounds which the author shows quite clearly to be a form of sonar by which the Cetacea are able to avoid collision when in the dark or in deep or muddy water. Clicks used by porpoises for this purpose are characterized by having the largest peak of acoustic energy in the sonic range although there are other vibrations far above the limits of audibility also present. There may be a series of peaks at between 80 and 120 kilocycles with smaller peaks at 140, 155 and 170 kilocycles. Evidence is also presented that the upper threshold of the auditory apparatus of porpoises is around 80 kilocycles. This is approaching the upper limit of hearing of the mouse (95 kilocycles) and the bat (98 kilocycles). There is much yet to be learned. This work in no way approaches the outstanding achievements of Griffin in analysing the sonar methods of bats. But it seems fairly clear that the animals can determine the distance and, within limits, the size of an obstacle from which its sounds are being reflected; and that these sounds are consistently used to avoid obstacles and to detect fish. It is clear that the porpoises must be highly expert at distinguishing the echoes from different kinds of surface and sea bottom. Perhaps the most intriguing fact is that all efforts to disturb or jam these echo-ranging signals by making extraneous interfering noises within the pool were failures. Even the reproduction of the porpoises' own sound pulses, which had been recorded as they swam among the test obstacles in earlier sessions, failed to cause collisions when played back to them through an underwater projector. It seemed obvious that they were able without difficulty to sort out the genuine echoes from the man-made copy.

W. H. Thorpe.

*Schizophrenia.* By F. J. Fish, M.B., M.R.C.P., D.P.M. Bristol: John Wright. 1962. Pp. xii + 190. 32s. 6d.

This book is offered to the postgraduate student of psychiatry and psychology as a textbook of moderate size and price. The author aimed to present those aspects of the subject which have been neglected in English-speaking countries, and to relate them to the views which are current in Anglo-American but especially British psychiatry. The result is a work of very individual stamp and of more than ordinary interest. Aspects of a theoretical kind, which in the usual textbooks are passed over with hardly a glance at the controversies they have caused, here receive extended treatment.

In Chapter III, for instance, we come on an account of the thought disorder of schizophrenia which extends over nearly 19 pages, with a fair and well-informed presentation of the theories of such workers as E. Bleuler, Cameron, Goldstein, Carl Schneider, Kleist, Jaspers, Conrad, Matussek and others. To practically all British readers at least one of these will be hardly more than a name—and for many, all of them. This section is of particular interest because we see, in all its nakedness, the search for what Dr. Fish has called "the philosopher's stone of schizophrenia, the basic psychological disorder from which all symptoms are derived." If only the psychologist and psychiatrist can remember that the primary change in schizophrenia is most probably of a chemical nature, he is unlikely to be seduced into looking for any unitary disorder of psychological function.

At the end of the book there is an equally educative chapter on the theories of schizophrenia, in which adequate space is given to psychoanalytic, Gestalt and existentialist schools. Material of this kind is very difficult for the British reader; that people should choose to think in such ways seems to him excessively outré. Dr. Fish succeeds, however, in providing a lucid account which seems to the reviewer both scholarly and well balanced.

ELIOT SLATER.

*Activation and Behaviour*. By Elizabeth Duffy. New York and London: Wiley. 1962. Pp. xvii + 384. 60s.

After re-reading the disarming and properly humble preface it is only with reluctance that a harsh judgement is passed on this book.

Unification is its stated aim. Dr. Duffy's intention is "to point out similarities in phenomena which might, without focus upon them, be overlooked," and "to present psychological phenomena in terms which will harmonize with those of other sciences, and to show the relationship of psychology to the foundation sciences upon which it rests." It is with her concept of energy mobilization, re-christened "activation," that this unification within psychology, and between psychology and physiology is attempted.

It is customary to use the term "activation" when referring to changes along a physiological dimension and "arousal" for the correlated behavioural changes. Dr. Duffy fails to maintain this distinction. "Activation" and "arousal" or sometimes "internal arousal" are used synonymously. The seeds of confusion thus sown, germinate in Chapter 6 where "internal arousal" and "intensity of behaviour" are equated, and come to fruition in Chapter 12 in which the main conclusion is stated, namely that activation (which through the mediation of "internal arousal" has been equated with "intensity of behaviour") is associated with important differences in behaviour. Even if the reader avoids this tautology by interpreting "activation" correctly, the conclusion is not altogether remarkable, and it was made available much more economically by Lindsley in Table I of a paper in *Electroenceph. Clin. Neurophysiol.*, 1952, **4**, 443–56 which Dr. Duffy herself quotes in a different context.

In the first section of the book (Chapters 1–6) the concept of activation is analysed. The second section (Chapters 7 and 8) concentrates on the sensory and motor effects of variations in activation. The last section (Chapters 9–11) is concerned with individual differences in activation. After studying these sections, which are well supplied with summaries, the reader may conclude that, by employing the concept of activation, unity is achieved. But he may question whether, with such a wide category as activation, this unification is of much more value than is the unification of the descriptive categories of psychology by ascribing them to the functioning of a nervous system. There is also the suspicion that Dr. Duffy believes that physiological explanations are better, in some absolute sense, than psychological explanations. If this is so she is on a slippery slope and can only provide needed ammunition for the empty organism brigade.

For those who take a jaundiced view of her thesis Dr. Duffy herself suggests that the eight or nine hundred references which she cites may prove useful "in the construction of a better image." That this card index catharsis will in fact be put to this use is doubtful. The serious student interested in a special field treated in the book will probably come away empty-handed, and those, who from the title expect an up-to-date summary of the work on the A.R.A.S. and the D.T.P.S., will be disappointed (or delighted) to find just a few relevant paragraphs which are principally derived from one secondary source.

It is with regret and a half a hope that I may be quite wrong that I conclude that Dr. Duffy's reputation was better served by her pioneering papers of the early nineteen-thirties than by this book.

J. R. SYMONS.

*Studies in Individual Differences: The Search for Intelligence*. Edited by J. J. Jenkins and D. G. Paterson. London: Methuen. 1962. Pp. xiv + 774. 63s.

This is a new volume in the Appleton-Century-Crofts Century Psychology Series and consists of 66 selected papers or extracts from larger works to which the Editors attach special historical importance. Among the authors represented are Galton, Spearman, Thorndike, Binet and Simon, Terman, Woodworth, Thurstone, Vernon, Anastasi and Guilford. The selection seems pretty sensible and the book should have its uses. It is perhaps a pity that nothing by Stern or Piaget has been included.

O. L. ZANGWILL.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XVI 1964 Part 2

# THE RUNNING MEMORY SPAN FOR WORDS

BY

LEO POSTMAN, THOMAS W. TURNAGE and ALBERT SILVERSTEIN
*From the University of California, Berkeley, California, U.S.A.*

This experiment investigates the running memory span for English words of high and of low frequency of usage. The running memory span is measured when a sequence of uncertain length is interrupted and the subject is required to recall the most recent part of the message. The span for the two classes of words was studied as a function of the length of the message and amount of prior practice. In the early stages of practice the span varied inversely with frequency of usage and increased with the length of the message. As practice continued, the difference between words of high and low frequency was largely eliminated and there was a relative decline in the span for the longest series. The major variables which are assumed to determine the length in the span are (*a*) the number and strength of pre-experimental associations of the items; (*b*) intraserial and interserial interferences developed during exposure to the series, and (*c*) the subjects' strategies of deliberate rehearsal.

## INTRODUCTION

The capacity for short-term storage of information is conventionally measured by the immediate memory span, i.e. the number of discrete units that can be reproduced in correct serial order after a single presentation. The standard procedure for measuring the span is to expose the subject to series of varying length and to require ordered reproduction of each message in its entirety. The definition of the span implied by these experimental operations has, however, limited generality. Short-term storage is not confined to discrete messages falling within the span of immediate memory. When a sequence in excess of the span is presented, some parts of the message can be reproduced correctly even though the entire sequence cannot. As the sequence progresses, the particular parts available for reproduction change continuously, i.e. old items drop out and new ones are added. As Waugh (1960) has suggested, the immediate memory span can be given a widely applicable definition as the length of a subseries within a longer series that can be recalled correctly after a single presentation. Such subseries are measured from reference points within the longer series. The beginning and the end of the series serve as the reference points for the determination of the initial and terminal span, respectively. By this definition the classical span based on the reproduction of the entire series represents the special case in which the initial and terminal spans coincide.

The position of the appropriate reference point in the determination of the immediate memory span depends on the segment of the total series which carries the information that is considered essential at the time of the test. The terminal span is critical when a continuous sequence is interrupted and the subject is required to recall the most recent parts of the message. This requirement obtains in practical situations in which questions about continuously changing events must be answered periodically. When the subject remains uncertain about the length of the series and hence about the point at which a test of recall will occur, the number of terminal items reproduced in the correct serial order defines the *running memory span* (RMS).

An extensive investigation of the major variables determining the RMS for decimal digits has been reported by Pollack, Johnson and Knaff (1959). The RMS was found to be consistently lower than the span obtained when the subject had advance information about the length of the series preceding the test. Both types of span showed some decreases with the length of the series but this decline was more pronounced for messages of uncertain than of known length. These differences were maintained over an extended period of practice. Pollack, Johnson and Knaff suggest that the accuracy of the terminal span and the differences between the results obtained with series of certain and uncertain length are a function of (*a*) the amounts of proactive inhibition falling on the items to be recalled, and (*b*) the "behavioural strategies" adopted by the subjects. The earlier parts of the message are assumed to interfere with the later parts, and the longer the total series the more opportunity there is for such interferences to develop. It is known that the amount of proactive interference increases as a function of the amount of prior learning (e.g. Keppel and Underwood, 1962; Underwood, 1945). The same relationship may be expected to hold for a continuous series, especially in the absence of distinctive cues for differentiating among successive parts of the message. The subject can, however, reduce the amount of effective interference by ignoring the early parts of the message. Such a strategy can be carried out more successfully when the length of the message is known than when it is uncertain. In the absence of information about the length of the series the subject may start active rehearsal of the terminal items either too early or too late.

It appears that the RMS provides a method for assessing rapidly accumulating proactive interference. Exploration of the conditions determining the RMS for materials other than digits, and especially verbal units, thus becomes of interest from the point of view of interference theory. There is considerable evidence that proactive inhibition plays a major part in determining the long-term course of forgetting for materials learned in the laboratory (Postman, 1961*a*; Underwood, 1957). For well practiced rote series proactive inhibition develops slowly during the retention interval (Postman, 1962*a*; Underwood, 1949) but recent findings show that such interference accumulates rapidly in the recall of verbal units to which the learner is given a single exposure (Keppel and Underwood, 1962; Murdock, 1961). It is reasonable to suppose, therefore, that the RMS will be sensitive to the conditions of interference among successive verbal units.

The present experiment compares the RMS for English words of high and low frequency of usage. The hypothesis was tested that susceptibility to inter-item interference would vary directly with the frequency of usage of the words; on this hypothesis, the RMS for words of low frequency should exceed that for words of high frequency. The higher the frequency of usage of a word the larger is the number of different verbal contexts in which it is likely to be used. The number and strength of pre-experimental associations acquired by a word should, therefore, vary with the frequency of usage. In agreement with this expectation, scaled indices of meaningfulness correlate highly with frequency of usage (Postman, 1961*b*; Underwood and Schulz, 1960). Pre-experimental associations are a source of both positive and negative transfer in the acquisition and retention of verbal sequences: they facilitate the development of associative links between successive units but they also compete with the prescribed sequences. The net transfer effect will depend on the degree to which pre-experimental and prescribed associations are compatible and on the opportunities for extinction and recovery of competing responses. The conditions under which the RMS is measured make it likely that interference will increase more rapidly than positive transfer as a function of frequency of usage. The series is presented only once so that there is little opportunity for unlearning erroneous associations. Furthermore, exact reproduction of the serial order is required, and it is known that the rate of misplaced responses increases sharply with the frequency of usage of the items (Postman, 1961*b*; 1962*b*).

The difference between the RMS for words of high and low frequency was investigated as a function of the length of the message and the amount of prior experience with the learning materials and the experimental conditions. In the light of the findings of Pollack, Johnson and Knaff (1959), these variables may be expected to influence the amounts of proactive inhibition falling on the terminal items and the development of "behavioural strategies" for reducing the amounts of effective interference.

## METHOD

### *Materials*

The stimulus items were two sets of 20 two-syllable nouns each. Set HF consisted of words of high frequency of usage classified as A or AA in the General Count of Thorndike and Lorge (1944). Set LF comprised words of relatively low frequency of usage; for these items the median frequency of occurrence in the General Count is 22, with a range from 10 to 39. (The General Count gives frequency of occurrences per million words. The frequency for words classified as A is between 50 and 100, for those classified as AA in excess of 100.) Norms of free association were available for the items. The frequency with which the members of a set elicited each other in the norms was zero in the large majority of cases and in no case exceeded 0·1 per cent. Since normative clusters were excluded, the possible differences in inter-item associations were limited essentially to those developed during exposure to the stimulus series.

Series of 10, 15 and 20 words were constructed from the words in each set. There were two different lists at each length. For the 20-item series, the two lists presented random sequences of all the words in a set. The two 15-item series had 10 words in common, and five words were unique to each list. There was no overlap between the words used in the two 10-item series. Thus, length and overlap between the two lists were correlated.

### *Design*

There were 12 experimental groups of 10 subjects each. Half the groups were tested with words from Set HF, and half the groups with words from Set LF. The groups were exposed to all six lists of the appropriate frequency, i.e. two lists at each of three lengths. A counterbalanced design was used so that each list was assigned once to each ordinal position in the sequence of six tests. For a given group, each length occurred once in one of the first three positions and once in one of the last three positions. Since each length was represented by two lists, every possible combination of length and ordinal position was given to 20 subjects.

### *Procedure*

The lists were recorded on tape and presented at the rate of one word per second. Two tones marked the end of each series. The subjects were provided with a record booklet and were instructed to start writing down as many of the terminal items as they could as soon as they heard the tones. The instructions made it clear that as many words as possible should be reproduced in the correct serial order, and that the reproductions would be scored with the last item as the point of reference. Since recall was written, the subjects were free to put the items down in any order they chose. Sixty sec. were allowed for recall, and there was an interval of 10 sec. between the end of a test and the start of the next series. A different page of the record booklet was used for each test.

### *Subjects*

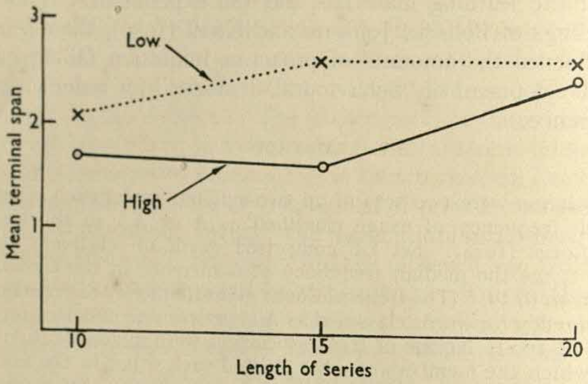
The subjects were 120 undergraduate students at the University of California. They were run in small groups averaging four in number. Assignment of successive groups to conditions was made by means of a table of random numbers, with the restriction that tests with Set HF and Set LF were alternated.

## RESULTS

### *Initial test series*

Performance on the first test of the experimental session was free from the effects of practice and transfer between successive lists. Thus the results of the initial test permit an assessment of the variables influencing the terminal span in the absence of any advance information about the composition or the length of the message. Figure 1 shows the mean RMS for words of high and low frequency of usage as a function of the length of the series. Each mean is based on the scores of 20 subjects. The span is significantly greater for Set LF than for Set HF ($F = 6{\cdot}57$, $d.f. = 1/114$, $0{\cdot}01 < p < 0{\cdot}02$). The RMS also increases reliably with the length of the message ($F = 3{\cdot}25$, $d.f. = 2/114$, $0{\cdot}02 < p < 0{\cdot}05$). The increases are negatively accelerated for Set LF and positively accelerated for Set HF, but the interaction of length and word frequency is not significant ($F = 1{\cdot}57$).

FIGURE 1



Mean RMS for first test series as a function of word frequency and length.

TABLE I

MEAN NUMBERS OF ITEMS REPRODUCED ON FIRST TEST AND RMS AS PER CENT. OF TOTAL RECALL

| *Length of series* | *Set HF* | | *Set LF* | |
|---|---|---|---|---|
| | *Mean No. items* | *Per cent. RMS* | *Mean No. items* | *Per cent. RMS* |
| 10 | 5·35 | 34·0 | 4·90 | 43·9 |
| 15 | 5·45 | 31·4 | 5·35 | 51·8 |
| 20 | 5·55 | 46·8 | 5·45 | 49·6 |

Table I shows the mean numbers of items reproduced on the test of recall without regard to serial order. These scores are substantially higher than the values of the RMS which are based on errorless reproductions of the serial order. The total numbers of responses tend to increase with the length of the series and to be larger for Set HF than Set LF. The latter finding indicates that the difference in RMS between the two sets of words is determined entirely by the accuracy of the serial order rather than by the availability of responses. This conclusion is clearly supported when the values of the RMS are expressed as percentages of the total amount recalled. These percentages are independent of the level of recall and reflect the subjects' ability to discriminate errors of serial position and to limit their reproductions to correct sequences. As Table I shows, the mean percentages increase as a function of the length of the series and decrease with word frequency. The differences between the two sets of words are, however, limited to the two shorter lengths. An analysis of variance of the percentages (following arc-sine transformation) shows word frequency to be a significant source of variance ($F = 5{\cdot}80$, $d.f. = 1/114$, $0{\cdot}02 < p < 0{\cdot}05$). The effects of length and the interaction of length with frequency fall short of significance ($F$'s of $2{\cdot}10$ and $2{\cdot}07$, respectively, $d.f. = 2/114$ for both).

*Practice effects*

In order to evaluate the influence of practice and transfer on the RMS, the results of the first and second half of the experimental session were compared for each length of series. As was indicated earlier, each length was used equally often in each ordinal position and for a given group appeared once in the first half and once in the second half of the session. Thus, the scores for the two halves, which are shown in Table II, represent the averages for the first three and last three ordinal positions, respectively.

TABLE II

MEAN RMS AND PER CENT. THAT RMS IS OF TOTAL RECALL FOR FIRST AND SECOND HALF OF EXPERIMENTAL SESSION

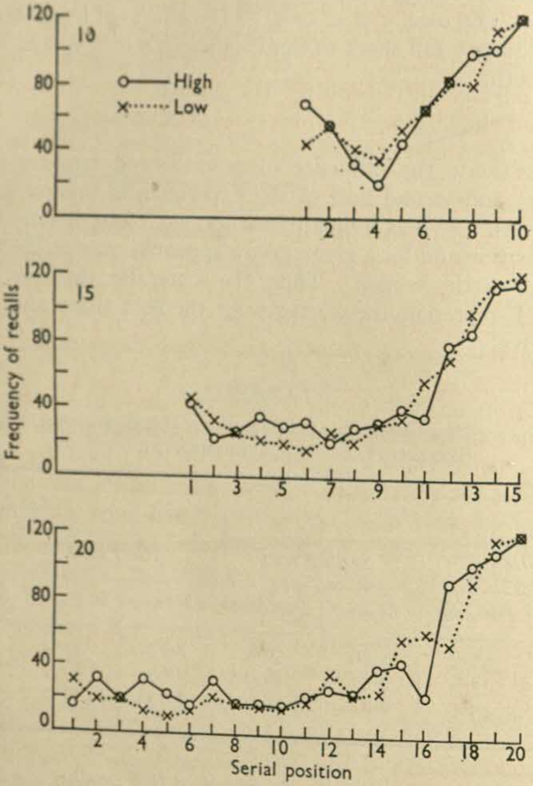
| | *Set HF* | | | | *Set LF* | | | |
|---|---|---|---|---|---|---|---|---|
| | *First half* | | *Second half* | | *First half* | | *Second half* | |
| *Length* | *Mean* | *Per cent.* | *Mean* | *Per cent.* | *Mean* | *Per cent.* | *Mean* | *Per cent.* |
| 10 | 2·00 | 38·5 | 2·80 | 50·8 | 2·17 | 41·0 | 2·58 | 46·1 |
| 15 | 2·05 | 37·8 | 2·82 | 46·9 | 2·67 | 49·5 | 3·07 | 52·7 |
| 20 | 2·50 | 45·4 | 2·72 | 41·6 | 2·73 | 49·4 | 2·57 | 41·6 |

The pattern of RMS scores for the first half of the session is essentially the same as that described for the initial series. Performance in the second half of the session was significantly better than in the first half ($F = 23{\cdot}50$, $d.f. = 1/118$, $p < 0{\cdot}001$). The amount of improvement was, however, substantially greater for Set HF than for Set LF, and also was almost entirely limited to the two shorter lengths of series. Order of presentation interacts significantly with word frequency ($F = 5{\cdot}10$, $d.f. = 1/118$, $0{\cdot}02 < p < 0{\cdot}05$) as well as with length ($F = 5{\cdot}18$, $d.f. = 2/118$, $p < 0{\cdot}01$). It may be noted in Table II that the changes between the first and second half follow essentially the same pattern when the RMS scores are expressed as percentages of the total amounts recalled. Thus, the difference between the two sets of words is eliminated with practice. In addition, there is a change in the relationship between length of series and RMS. Whereas initially the RMS increased with length, the longest series showed a disadvantage in the later stages of practice. This conforms to the results for digits obtained by Pollack, Johnson and Knaff (1959) with practiced subjects.

### *Serial-position curves in recall*

While the RMS increased with practice, the distribution over serial positions of the responses outside the span did not change systematically. The results obtained at all stages of testing were, therefore, combined for purposes of plotting the serial-position curve for series of a given word frequency and length. Each of the curves shown in Figure 2 is based on 120 cases (60 subjects tested twice). The frequencies of recall, which are plotted against serial position during presentation, include the terminal span as well as responses falling outside the span. For the 10-item series a primacy as well as a recency effect is obtained. In spite of their greater remoteness from the terminal point of reference, items at the beginning of the list are more likely to appear in the recall protocols than those from the middle of the list. The curve is more steeply bowed for Set HF than for Set LF. The primacy effect is substantially reduced for the 15-item series and disappears altogether for the 20-item series. The decrease in the primacy effect with length obtains for both types of words.

FIGURE 2



Serial-position curves at recall as a function of word frequency and length.

Words remote from the end of the series which are reproduced frequently during the recall of the terminal items are likely to represent highly effective sources of interference with the RMS. Figure 2 indicates that the amount of such competition from the initial segment of the list is inversely related to the length of the series. Since the RMS is of the order of two to three words and in no individual case exceeded seven words, recall of the first three items never contributed to the terminal span.

### *Intrusions*

Successive series were drawn from the same pool of words, but only a part of the total pool was represented in each of the 10-item and 15-item series. Starting with the second test in the experimental session, overt interlist intrusions could, therefore, occur during the recall of these series. Such intrusions could not be identified in the recall of the 20-item series which comprised the entire pool of items. Tests on the 10-item series yielded a total of 46 interlist intrusions for Set HF and 39 such errors for Set LF. The corresponding totals on tests of the 15-item series were 27 and 19. The number of possible intrusions for the 10-item series was twice that for the 15-item series, and the observed numbers of occurrences closely reflect the ratio of opportunities. In both cases the number of intrusions is slightly larger for Set HF than Set LF.

To the extent that the same items occur in successive lists, intraserial errors and intrusions cannot be distinguished. It is known, however, that there is massive interlist interference when different serial orders of the same items are learned consecutively (Irion, 1946). Hence the reduction in identifiable intrusions does not in any way imply a decrease in interlist interference as a function of length. On the contrary, such interference should increase with the number of elements a given test list shares with prior lists.

Importations from outside the list occurred throughout the experimental session. As subjects became increasingly familiar with the items in the pool the number of importations declined. The total number of importations during the first half of the session was 39 for Set HF and 36 for Set LF. In the second half of the session the corresponding numbers were 14 and 15. There were only minor variations in the number of importations as a function of length of list.

## DISCUSSION

When a subject has no information about the composition or probable length of a message the RMS varies inversely with the word frequency of the items and shows some improvement with the length of the series. It was suggested earlier that the strength of inter-item associations increases with frequency of usage and that competition from such associations interferes with the reproduction of the prescribed serial order. The results obtained on the initial test series are in agreement with this expectation. Since total recall was slightly greater, and the per cent. RMS clearly smaller, for Set HF than for Set LF, word frequency appears to have a dual effect—it increases the probability of recall but at the same time disturbs the reproduction of serial order. It should be emphasized that there were no normative associative connections among the members of either set. The inter-item associations responsible for the differences in performance were, therefore, established during exposure to the stimulus series. It is likely that such associations reflect the development of stable mediational links.

The finding that word frequency has different effects on amount recalled and on RMS shows that the influence of language habits on retention varies with the criterion of performance. The presence of inter-item associations favours recall since members of the series are likely to elicit each other during the test of retention (Deese, 1959). At the same time, inter-item associations interfere with the reproduction of serial order because there is a high probability that the succession of such associations during the test does not conform to the prescribed arbitrary sequence. Parallel results have been obtained in studies of retention as a function of inter-item similarity. As similarity increases, there is a progressive restriction in the range of alternative responses which favours recall. Reduced discriminability of individual items leads, however, to errors of generalization and disruption of associative links (Horowitz, 1961; Underwood, Runquist, and Schulz, 1959). Thus, conditions which raise the availability of a class of responses are also likely to generate intraserial interferences. The RMS which is critically dependent on the accuracy of serial order is highly sensitive to such interferences.

Two factors may contribute to the increase in RMS with length of list in the initial stage of the experiment. First, as Figure 2 shows, the shorter the message the greater is the amount of interference from the relatively strong items at the beginning of the series. As length increases there is a reduction in the number of items of comparable competitive strength. After a single exposure, the strength of individual competing responses may have greater weight than the sheer number of potential competitors. Second, it appears probable that as a message of uncertain length continues subjects become increasingly set for the occurrence of the test and begin rehearsing the most recent items. The fact that increases in RMS with length occur more slowly for Set HF than Set LF suggests that rehearsal may be less effective in the former case, perhaps because of a greater likelihood of erroneous as well as correct associations occurring in rehearsal.

As practice continues, performance on the two sets of words becomes equally proficient. Improved differentiation among the items in Set HF may be responsible for the elimination of the initial difference between the two kinds of materials. In successive series subjects are exposed to different sequential arrangements of items from the same pool and learn to discriminate among them. In agreement with this interpretation, the percentages of RMS to the base of total amount recalled are essentially equal for Sets HF and LF in the second half of the experiment. As was suggested earlier, these percentages reflect subjects' ability to discriminate correct sequences and to reject incorrect ones.

The positive relationship between length of series and RMS is also largely eliminated with practice. In the second half of the experiment performance actually tends to drop as the length of the series is increased from 15 to 20 items. It is perhaps not surprising that the factors to which the initial positive relationship between length and RMS was attributed lose their effectiveness. With repeated exposures to items from the same pool interserial interferences come to mask intraserial effects. The longer a series the more items it shares with each of the preceding ones and hence the greater the probability that the terminal items will be subject to interference from prior series. As for the increased likelihood of rehearsal toward the end of a long series, the subjects gain experience with the range of possible lengths and hence may be assumed to start rehearsal earlier as practice continues. Early rehearsal is favourable to the reproduction of short series and at the same time increases the strength of responses which will compete with the terminal items of long series.

It is clear that the practice effects which we have examined are specific to a situation in which a limited pool of items is used and the overlap between successive series is correlated with length. The same situation should obtain on the average in the measurement of the RMS for decimal digits. The relationships observed in the initial test would be expected to persist over longer periods of practice if new materials continued to be introduced.

A final comment is in order about the shortness of the RMS for words which is less than half as long as the RMS for digits (Pollack, Johnson and Knaff, 1959). Digits also have an advantage over words when the memory span is measured by the classical procedure (Crannell and Parrish, 1957) but the difference is not nearly as great. In general, the difference between the spans for the two types of materials may be related to the subjects' pre-experimental experience in dealing with arbitrary strings of digits as compared with arbitrary sequences of words. Different serial arrangements of digits are in constant use, and each subject may be presumed to have a vast amount of experience in differentiating among such sequences and developing habits for "encoding" them. Sequences of unrelated single words are encountered and recalled much less frequently. Thus a chunk of information (Miller, 1956) should include more digits than unrelated words, even after some experimental practice. Such transfer effects of pre-experimental habits gain in importance as the amount of interference at recall increases. Hence the difference between digits and words is enhanced in the RMS as compared with the classical memory span.

This research was carried out under Contract Nonr 222(90) between the Office of Naval Research and the University of California. The work was performed at the Institute of Human Learning which is supported by a grant from the National Science Foundation.

## REFERENCES

CRANNELL, C. W., and PARRISH, J. M. (1957). A comparison of immediate memory span for digits, letters and words. *J. Psychol.*, **44**, 319–27.

DEESE, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, **5**, 305–12.

HOROWITZ, L. M. (1961). Free recall and ordering of trigrams. *J. exp. Psychol.*, **62**, 51–7.

IRION, A. L. (1946). Retroactive inhibition as a function of the relative serial positions of the original and interpolated items. *J. exp. Psychol.*, **36**, 262–70.

KEPPEL, G., and UNDERWOOD, B. J. (1962). Proactive inhibition in short-term retention of single items. *J. verb. Learng. verb. Behav.*, **1**, 153–61.

MILLER, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, **63**, 81–97.

MURDOCK, B. B., Jr. (1961). The retention of individual items. *J. exp. Psychol.*, **62**, 618–25.

POLLACK, I., JOHNSON, L. B., and KNAFF, P. R. (1959). Running memory span. *J. exp. Psychol.*, **57**, 137–46.

POSTMAN, L. (1961*a*). The present status of interference theory. In C. N. COFER (Ed.), *Verbal Learning and Verbal Behavior*. New York.

POSTMAN, L. (1961*b*). Extra-experimental interference and the retention of words. *J. exp. Psychol.*, **61**, 97–110.

POSTMAN, L. (1962*a*). The temporal course of proactive inhibition for serial lists. *J. exp. Psychol.*, **63**, 361–9.

POSTMAN, L. (1962*b*). The effects of language habits on the acquisition and retention of verbal associations. *J. exp. Psychol.*, **64**, 7–19.

THORNDIKE, E. L., and LORGE, I. (1944). *The Teachers Wordbook of 30,000 Words*. New York.

UNDERWOOD, B. J. (1945). The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 59, No. 3.

UNDERWOOD, B. J. (1949). Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, **39**, 24–34.

UNDERWOOD, B. J. (1957). Interference and forgetting. *Psychol. Rev.*, **64**, 49–60.

UNDERWOOD, B. J., RUNQUIST, W. N., and SCHULZ, R. W. (1959). Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, **58**, 70–8.

UNDERWOOD, B. J., and SCHULZ, R. W. (1960). *Meaningfulness and Verbal Learning*. Philadelphia.

WAUGH, N. C. (1960). Serial position and the memory-span. *Amer. J. Psychol.*, **73**, 68–79.

*Manuscript received* 28*th August*, 1962.

# THE ORIGIN OF THE AFTER-EFFECT OF MOVEMENT

BY

MARY J. PICKERSGILL* and M. A. JEEVES†

*From the Departments of Psychiatry and Psychology, University of Leeds*

Previous experimental work on the origin of the after-effect of seen movement is discussed and the relevance of the argument from transfer, which is generally used to demonstrate that the effect is central, is brought into question. The problem is reformulated in terms of whether a retinal contribution is necessary throughout the after-effect and experiments are described which support the hypothesis that it is necessary. It is suggested that perception of the movement after-effect also involves normal movement perception, which itself is dependent on the integrity of cerebral functioning.

## INTRODUCTION

It may be accepted that any perception involves central factors. Retinal stimulation is also generally involved in visual perception; where it is not, it is usually appropriate to talk of hallucinations. After-effects, however, are a special case: something is being seen in the absence of external stimulation, but, unlike an hallucination, the nature of the perception depends upon the nature of the immediately preceding stimulation.

If it is true that central processes are taking place throughout, then the important question which remains concerning the origin of the after-effect of seen movement is whether retinal ones are also present throughout. Hitherto, however, the question has not been posed in just these terms and a great variety of experiments has been performed. The variety of explanations of its occurrence that has been proposed is only slightly less extensive, but a broad distinction may be made between those laying greatest emphasis on central processes and those on peripheral ones.

Several experimental approaches have been used in the past and the findings from these will first be summarized.

### (*a*) *Experiments on the effects of antagonistic stimulation of* (i) *different parts of the retina, and* (ii) *corresponding retinal points*

(i) Gates (1934) using a rotating spiral, and Wohlgemuth (1911) and Durup (1928), who used mainly the "waterfall illusion," are agreed that opposite stimulations to the same retina result in two opposed after-effects. This sheds no light on the question whether processes underlying the after-effect are occurring in the retina.

(ii) Wohlgemuth and Durup find that equal and opposing movements exciting corresponding retinal points result in no after-effect, or a very faint one, with binocular vision, although with monocular vision the expected after-effect for that eye is seen. This implies (1) there is interaction between the two effects at some point beyond the retina, (2) that *the effects continue separately* at some level below this so that perception on the monocular effect is still possible and (3) that closing one eye prevents its after-effect and the subsequent interaction.

The situation is further complicated by Hunter's (1914) report that opposing after-effects *can* be seen in corresponding retinal areas. He used a prism to superimpose the opposite stimulations presented to the right and the left eyes. Gates showed two opposing after-effects in conditions in which opposite stimulations were presented to each eye separately. It is difficult to be certain in these experiments how accurately stimulation was confined to corresponding points.

---

* Now at Department of Psychology, Bedford College, University of London.

† Now at Department of Psychology, University of Adelaide.

As with a single retina, a result in this direction, i.e. the occurrence of two opposing after-effects, sheds no light on whether these are dependent on central and retinal, or only central processes.

### (*b*) *Experiments comparing monocular and binocular effects*

Durup noticed a marked diminution in length of after-effect if the after-effect was observed monocularly after binocular stimulation, and a much greater one with monocular fixation and binocular observation of the after-effect. This suggests interaction between the visual systems connecting the two eyes. It is not clear, however, whether the diminution of after-effect in Durup's experiments is due to competitive effects, or whether processes resulting from stimulation of both eyes normally have an additive effect. It was decided to make a comparison between the effects of using one eye and of using both eyes throughout both stimulation and observation of the after-effect.

### (*c*) *Experiments on "transfer" to the unstimulated eye*

Isaac Newton, in a letter published in 1831, describes how an after-effect resulting from looking at the sun with his right eye could be seen with his left eye, "almost as plain as with my right eye, if I did but intend my fancy a little while upon it."

Observation of the after-effect of seen movement with the unstimulated eye after monocular stimulation has been reported by Wohlgemuth, Hunter, Durup and Holland (1957). Hunter obtained the effect with the waterfall illusion only, and not with the spiral, although he was obliged to persuade his subjects that the "disturbance" they saw after stimulation by the spiral was not an after-effect. It is generally noted that the "transferred" after-effect is fainter and Holland finds the average duration is shorter. Holland (1957) considers that in so far as the after-effect does not transfer it is retinal, but concludes from the amount of transfer found that "little doubt can now be held concerning the central nature of the after-effect" (1960, p. 203). Hunter argues similarly, but since he finds total lack of transfer he concludes this gives "emphatic support to the retinal character of the after-movement secured from Plateau's spiral" (Hunter, 1914, p. 259). Thus both consider the report of transfer to provide evidence from which the relative contributions of retinal and central processes can be deduced. Wohlgemuth and Durup, on the contrary, are of the opinion that transfer could be explained by the projection of the process taking place in the closed eye on to the field of the open eye.

Thus there is disagreement on two counts: first as to whether "transfer" takes place after stimulation by movement, and secondly as to its implications.

As previous workers had used very few subjects it was decided to discover the frequency and duration of "transferred" after-effects in a larger group.

That the implications of transfer are unclear is apparent from the different inferences drawn by different workers. The very use of the word begs the question; the fact is that the after-effect can be seen with the unstimulated eye open, and the stimulated one closed. To say that "transfer" has occurred is to give an interpretation to this fact, and one whose meaning is not at all clear. It will be argued here that the report of transfer does not demonstrate anything about the origin of the after-effect. Further examination shows that it is very difficult to give any satisfactory meaning to the concept if it is to be used in an *explanatory* way to refer to a process, and not simply as a question-begging description.

Possible meanings of "transfer" are:

(1) *That the after-effect has literally transferred to the opposite retina.* This finds possible support in recent reports that centrifugal fibres may be present in the human optic nerve (Granit, 1955; Polyak, 1941). However, even if such fibres exist, their presence would not establish whether or not processes on the *stimulated* retina were a necessary condition of the after-effect and thus the locus of the processes which are to be transferred would still be unclear.

(2) *That the after-effect can be seen "with" the unstimulated eye.* In normal usage, "Which eye are you seeing with?" would be equivalent to, "Which eye is open?" because seeing normally implies perception of the external world, and this is known to be impossible when the eye is shut. In the case of after-effects this equivalence breaks down, since perception is in fact taking place in the absence of external stimuli, and the question, "Which eye is seeing the after-effect?" needs a reinterpretation of the word "seeing." One might be asking *which eye received the stimulus* inducing the after-effect, but in that case it would be false by definition to assert that it could be seen with the other eye, and nonsense to enquire whether transfer had occurred.

(3) *That the process taking place after monocular stimulation does not convey sufficient information to allow discrimination of which eye it originated in.* If it does not, it is not very meaningful to say which eye the effect is "in"; however, since it can be seen with either eye open, one might feel tempted to say it is "in" both, and that "transfer" had occurred. But the use of the word "transfer" in this context is not very appropriate if the effect is not actually being discriminated by the use of the other eye, i.e. if the report would be the same if that eye were removed. Restricted only to instances where discrimination breaks down, the term "transfer" seems inappropriate and meaningless.

Proponents of the argument from transfer use it to try to show that an after-effect is "central." However, this is obviously true; if the optic nerve is severed, perception does not take place at all, and to this extent *all* perception is central. It is possible to ask whether retinal processes are *also* necessary during the after-effect as a condition of its perception, but this is a separate question, and it has been shown that the occurrence of transfer, i.e. the willingness to ascribe the effect to the unstimulated eye, is not relevant to its solution.

### *Explanation of reports of transfer*

There is evidence that the explanation of this willingness to ascribe the after-effect to the unstimulated eye is that discrimination of which eye is giving or has given rise to an after-effect is not possible, and since perception is present, and one eye is open, the influence of past experience is to lead the subject to say he is seeing the effect "with" the open eye. If discrimination were possible, then the report of transfer would be much more difficult to explain. The results of an experiment by Pickersgill (1961) support the hypothesis that in the absence of the usual cues giving information about whether an eye is open or shut, and the characteristic shape of its visual field, it is impossible to tell which eye is being stimulated.

There is some evidence (Oswald, 1957, quoting Cibis and Nothdurft, 1948; Craik, 1940; Brindley, 1959) that the after-effect of light depends normally on retinal processes which are at least partly photochemical in nature. After-images to hallucinated stimuli have been reported (Oswald, 1957) but this is hardly likely to be the normal mechanism of their occurrence. If one accepts this evidence, then the report of an after-effect "in" the unstimulated eye can only be due to the subject's not knowing which eye it is in. Pickersgill (1959) showed that subjects do report "transfer" of the after-effect of light from a normal eye, and also from an eye rendered pressure blind during stimulation, thus providing further evidence of the inability to discriminate.

### *Experimental techniques*

As the report of transfer is not very informative as to the origin of the after-effect of movement, other approaches to the problem must be made.

Craik (1940) followed up an observation by Helmholtz (1924, p. 7) that pressure on the eye causes blindness due to embarrassment of the retinal blood supply and subsequent anoxia, and suggested that "this blindness affords a convenient method of blocking the transmission of nerve impulses between the eye and the brain.". His demonstration that an after-effect of light could be produced in a pressure blind eye he thought confirmed its retinal and photochemical origin.

If the after-effect of movement were *absent* after anoxia during stimulation, this would suggest *either* that cerebral participation during stimulation is necessary, *or* that the after-effect of movement is dependent on neural processes in the retina and that the effect of anoxia is to prevent an adequate response in the retina. To decide between these two possibilities it would be necessary to discover whether an interference with retinal processes *after* stimulation can terminate the after-effect. If it is *only cerebral* it should continue to be visible to the unstimulated eye while the stimulated eye is rendered anoxic; also, once it has ended, release of pressure in the stimulated eye should not be followed by any further after-effect with the return of vision.

One problem about this technique is the difficulty of knowing whether the anoxia is sufficient to prevent *all* transmission along the optic nerve. The fact that one is no longer aware of the contents of the visual field is no guarantee of this. In fact Craik and Vernon (1941) investigated the relationship between amount of pressure and onset of blindness. They found that pressure causes a rise in threshold for light commencing after 5–15 sec. and increasing for 30–40 sec. The minimum pressure to produce any loss of vision (about 80 mm. Hg.) is considerably less than that found necessary by Duke-Elder (1926) to produce complete obliteration of the retinal arteries (about 115 mm. Hg.)

The likelihood that vascular occlusion is only partial was confirmed by ophthalmoscopic examination of the fundi of a few subjects during pressure on the eye (Pickersgill, 1959). Pressure resulted in almost immediate constriction and blanching of the vessels at the points of entry and exit, which was followed by dilation of the veins above the area of constriction immediately surrounding the papilla, and very marked pulsation of the artery. This was seen to continue for up to 40 sec.; complete collapse of the circulation was never seen and the veins remained filled. Subjects usually reported momentary return of vision at some point during examination.

In the ensuing discussion we shall for convenience refer to an eye made temporarily pressure blind in this way as anoxic, although it was clearly impossible to establish in each instance the precise extent of the anoxia. This does not affect the argument provided a sufficient degree of anoxia can be achieved to produce an effect under the conditions described above.

A modification of Craik's (1940) technique was found to be necessary. He used the normal eye to control fixation of the anoxic eye and it therefore received the same stimulation. It could be argued that this procedure invalidates his own result as an after-effect in the normal eye could have been "seen" in the anoxic eye. His finding is in fact confirmed by Cibis and Northdurft (1948), who stimulated the blind eye only; however, in these present experiments a technique was used which allowed the normal eye to control fixation, but prevented it from being stimulated in the same way as the anoxic eye.

## METHODS AND PROCEDURE

### *Apparatus*

Only the after-effect of movement produced by a rotating spiral has been investigated, and the same apparatus was used throughout the experiments. This consisted of a Goldring Lenco (GL55) gramophone motor, which had the usual four fixed speeds, 16, 33, 45 and 78 r.p.m., and was capable of continuously variable adjustment within this range. The speeds were guaranteed by the makers not to vary more than 1 per cent., but only if the motor was kept horizontal. It was therefore built into a box in this position, and a mirror placed on the underside of the lid. When the lid was 45° open, a vertical image of the horizontal turntable was seen in the mirror. The apparatus was painted black apart from the turntable, which was greyish in colour. Between the outer edge of the box and the lid was placed a cardboard screen in which was cut a circular hole 7 in. in diameter; this prevented the subject from seeing any reflection in the mirror but that of the spiral and the surrounding edge of the turntable. The centre of the turntable was 12 in. from the front of the box.

Unless otherwise stated, the pattern consisted of four equiangular spirals each making 1¼ revolutions, which appeared to unwind increasingly as they left the centre.

The apparatus was illuminated by two jointed "Pifco" lamps, holding 60-watt bulbs, one being placed on each side and directed on to the turntable. In the experiments in a darkened room, distracting stray light was eliminated by draping with a black cloth.

The precise modifications of the apparatus needed for each experiment will be described separately.

### *Subjects*

The subjects for these experiments were almost entirely students and staff of Leeds University. The majority were in their early twenties, and most of the staff were in their twenties or thirties.

### *Instructions to subjects*

The instructions, which were given orally, were "I want you to look very carefully at the centre of the black and white spiral in front of you. When I say 'Ready,' I shall switch on and the disc will go round. After about half a minute I shall stop it, and I want you *to continue to watch the centre*. You may see some kind of after-effect, and if you do I want you to tell me when it has ended, that is, when the pattern looks normal again, as it is now. If, on the other hand, the spiral appears normal as soon as it stops, just say so."

Subjects were always asked whether they understood, and an attempt was made to clear up misunderstandings before testing began. The full instructions were not repeated to subjects who participated in more than one experiment.

Intervals between trials were not standardized, mainly because an unduly long interval would have to have been selected in order to make sure that delays resulting from, e.g. a long verbal report by the subject, would fall within the interval allowed.

## EXPERIMENT I

A comparison of lengths of after-effect resulting from

(i) monocular stimulation and observation with the stimulated eye;
(ii) binocular stimulation and observation with both eyes;
(iii) monocular stimulation and observation with the unstimulated eye.

### *Method and procedure*

The subjects were seated about 5 ft. from the front of the box, looking through an eye-piece. A 30 sec. stimulation at 45 r.p.m. was used, and the room was dark except for the illumination of the spiral by the two lamps. Under conditions one and two the after-effect was observed with the eye or eyes that had been stimulated. With condition three, after monocular stimulation the stimulated eye was closed and the other opened. There were two trials, one in each direction, i.e. from left to right and right to left.

*Results*

The mean lengths of after-effect on all three conditions are shown in Table I.

TABLE I

MEAN LENGTHS OF AFTER-EFFECT FOLLOWING 30 SEC. STIMULATION

| *Stimulated eye* | *Eye observing the after-effect* | *N* | *Mean length of after-effect in seconds* |
|---|---|---|---|
| R | L | 25 | 7·1 |
| L | R | 25 | 6·6 |
| R | R | 19 | 11·6 |
| L | L | 19 | 11·3 |
| R + L | R + L | 19 | 13·9 |

A striking overall diminution in length of after-effect in the unstimulated eye is apparent. Perhaps more important, 11 of the 25 subjects noticed no effect at all in the unstimulated eye, and 5 in one direction only. If the results of only those subjects who did find transfer are considered, then the duration still tends to be less, but not to a marked extent and not in all subjects.

### EXPERIMENT II

To determine whether the after-effect of movement will occur if the eye is made pressure blind during stimulation.

*Method and procedure*

Subjects were shown the technique of pressing on the outer part of the eye-ball, over the lid, in such a way as to produce loss of vision and to keep the pupil unobscured by the lid. With a little practice most subjects were able to achieve this condition without any difficulty; they were asked to release their eye if they experienced any discomfort.

The other eye was used to control fixation, and at the same time protected from the stimulus to the pressure-blind eye by a cardboard screen in which there was a pinhole about 1 in. in from the nasal edge. The subjects were asked to hold this in front of the "non-experimental" eye and adjust it so that only the central fixation knob could be seen through it. They then closed this eye and opened the other to make sure that the screen was not obscuring its view of the disc.

Several subjects found that under these conditions two images were seen, one with the "non-experimental" eye, through the pinhole, and one with the "experimental" eye, resulting in the perception of two separate centres; with slight alterations of the position of the screen they usually managed to fuse the centres or adjust them so that they did not appear more than about an inch apart.

When this adjustment had been made and the screen was in place the subject began to press on the unscreened eye and told the experimenter when he could no longer see through it. A four-spoke 9 in. expanding spiral was rotated for 30 sec. at 45 r.p.m. Nine subjects were tested in both eyes.

*Results*

No after-effect of movement at all, on either trial, was seen by 8 of the 9 subjects. One subject reported a slight after-effect, but on questioning it was discovered that she had been aware of the stimulus as "movement," although not of its details, i.e. blindness had not been complete. Presumably subjects had been "fixating" as all reported seeing the spiral before anything else, and within 2 or 3 sec. of releasing their eye-ball. All subjects had previously reported seeing an after-effect under normal viewing conditions and with the same stimulation period.

It would appear at first sight that the after-effect of movement, unlike that of light, is dependent on impulses arriving at the cortex during stimulation, and therefore that cerebral participation during stimulation is necessary for the perception of the after-effect.

However, it is also possible that the effect of the anoxia induced by pressure is to prevent the retinal elements responding adequately. If it is severe enough to block synaptic transmission and produce blindness, then it may well block neural activity on the retina. The after-effect of *light* would still occur in the anoxic eye because it depends primarily on a photochemical process which is not affected by lack of oxygen. Pressure may be applied to the eye after instead of during stimulation. This would allow the initial stimulus to reach the higher centres in the usual manner. If this is *all* that is needed to produce an after-effect, pressure on the eye *after* stimulation should not have any disturbing effect. One could then assume that the result of the preceding experiment was due to failure of the message to reach the higher centres. If, on the other hand, pressure after stimulation does interfere with the after-effect, the possibilities are:

(*a*) retinal processes are essential throughout and they can be upset by anoxia at any stage of stimulation or after-effect;

(*b*) retinal processes are not necessary but stimulation of the eye by pressure causes interference with the central processes.

## EXPERIMENT III

To determine whether retinal anoxia induced after stimulation of the eye would destroy or attenuate the after-effect of movement in that eye.

### *Method and procedure*

In order that any effects of anoxia should be manifest, it was necessary it should be induced and recovery completed in a period shorter than that of the normal after-effect for that subject. An attempt was therefore made to select subjects whose length of after-effect was at least 15 sec. and 11 took part in the full experiment. A four-spoke 9 in. expanding spiral was used, and the subject sat about 3 ft. 6 in. from the apparatus. Stimulation was for 30 sec. at 45 r.p.m.

In order to demonstrate the after-effect, it was first measured using both eyes. One eye alone was then stimulated and any changes in the after-effect in that eye were observed under each of three post-stimulation conditions:

(1) normal observation of the disc with the stimulated eye;

(2) closing the eye for 10 sec. after stimulation;

(3) pressing on the eye until blindness was reported and maintaining the pressure until asked to release the eye.

Under all conditions the subject was asked to report whether he saw an after-effect, and, if so, when it ended. Condition 2 was the control for 3 as it was found that blindness was usually reported within 10 sec. The opposite eye was kept covered throughout.

### *Results*

All lengths of after-effect reported in Table II refer to the time from the end of stimulation to the end of the after-effect on the first trial in each condition.

The subjects all took part in more than one trial, and the number of trials with Condition 3 varied from 2 to 9, depending on the subjects' expressions of fatigue and discomfort. In a few subjects both eyes were tested. Conditions 1 and 2 were not always repeated after each trial of Condition 3 if they were giving consistent results.

## TABLE II

Time to end of after-effect in seconds on first trial under three conditions:

(1) normal observation with the stimulated eye;

(2) closing the eye for 10 sec. after stimulation;

(3) pressing on the eye-ball after stimulation.

The periods during which pressure is initiated and maintained are also shown for each subject. N = 11.

| | *Subject* | *A* *Period of pressure before blindness reported* (*sec.*) | *B* *Period blindness maintained* (*sec.*) | *A + B* *Total period of pressure* (*sec.*) | *Length of after-effect on first trial* (*sec.*) | | |
|---|---|---|---|---|---|---|---|
| | | | | | *Condition* | | |
| | | | | | 1 | 2 | 3 |
| Group 1 reporting no after-effect under Condition 3 | 1 | 5 | 1 | 6 | 15 | 23 | 0 |
| | 3 | 7 | 3 | 10 | 16 | 19 | 0 |
| | 4 | 12 | 4 | 16 | 20 | 31 | 0 |
| | 7 | 12 | 0 | 12 | 24 | 25 | 0 |
| | 8 | 11 | 0 | 11 | 16 | 20 | 0 |
| | 9 | 9 | 0 | 9 | 17 | 17 | 0 |
| | Mean | 9·3 | 1·3 | 10·7 | 18·0 | 22·5 | 0 |
| Group 2 reporting after-effect under Condition 3 | 2 | 10 | 5 | 15 | 22 | 19 | 27 |
| | 5 | 5 | 10 | 15 | 19 | 23 | 28 |
| | 6 | 12 | 3 | 15 | 24 | 26 | 38 |
| | 10 | 9 | 0 | 9 | 12 | 20 | 25 |
| | 11 | 7 | 3 | 10 | 17 | 19 | 23 |
| | Mean | 8·6 | 4·2 | 12·8 | 18·8 | 21·4 | 28·2 |

*Note:* On this first trial all the subjects released their eye after a period of pressure blindness which was less than the length of after-effect in both Condition 1 and Condition 2.

Six subjects saw no after-effect on the first trial of Condition 3. Further:

(i) Five saw no after-effect on any subsequent trial. One subject (No. 3) thought he did so on one occasion but was doubtful.

(ii) The total time during which pressure blindness was achieved and maintained was on all trials for one subject less than any length of after-effect for that subject on all trials under Condition 2.

(iii) For 4 out of the 6 subjects the total time during which pressure blindness was achieved and maintained was on all trials for one subject less than any length of after-effect for that subject using the same eye in Condition 1. For the other 2 subjects (Nos. 1 and 4) the maximum period of blindness was less than the minimum length of after-effect in Condition 1 on 3 out of 5 trials.

It would, therefore, be true to say of all these subjects that there should have been adequate time after the release of the pressure blind eye for an after-effect to have been experienced on some of the trials.

The other 5 subjects, who saw after-effects under Condition 3, reported the movement as being much fainter and harder to discriminate. Furthermore, some rather puzzling but nevertheless consistent results were found. It was noticed that if an after-effect did occur after pressure, it would frequently continue until well after the normal after-effect would have ended. Further trials were therefore undertaken with some subjects to discover for how long pressure had to be maintained for the after-effect not to reappear. This point varied, but for some subjects pressure could be maintained after the end of the normal after-effect, and a further 10–15 sec. of after-effect would be seen, suggesting some kind of lengthening effect.

This was very marked in subject 11 who claimed a definite after-effect of 9 sec. even when the eye was released 35 sec. after stimulation. However, when he was retested later, this effect did not occur.

Closing the eyes for 10 sec. after stimulation also postpones the ending of the after-effect, but not to such a marked extent. There is a mean rise of 3·6 sec. between the first trials on Conditions 1 and 2. This result is significant at the 2 per cent level of confidence.

There is good evidence that retinal anoxia after stimulation may destroy the after-effect of movement. Two possible explanations have already been suggested in the discussion of Experiment II: (*a*) the after-effect is dependent on retinal and neural processes, as seems most likely. The fact that this does not occur in all subjects is not crucial. There are obviously considerable differences in the degree of pressure exerted and in the resulting oxygen deprivation; the different amounts of time needed to report blindness by different subjects and by the same subject on different occasions are evidence of this. A possible objection would be that synaptic transmission is not normal for a few sec. after pressure has stopped and during this time the normal after-effect would have come to an end. This is not very plausible as vision for contours is restored almost immediately. (*b*) The other possible explanation of the loss of the after-effect with pressure after stimulation is that it is entirely dependent on central processes, but that the pressure in some way upsets the eye and the ability to see the after-effect projected on to a background. If that is so, observation with the eye which has *not* been submitted to pressure should remain unhampered both during and after pressure on the unstimulated eye. This is investigated in the following experiment.

It seems paradoxical that if the after-effect is not destroyed it is potentiated, at least in duration; in appearance it is much fainter. The facts that closing the eyes after stimulation has a similar effect, and that the after-effect is very difficult to see with the eyes closed could be taken to mean that the adaptation resulting from stimulation will not be apparent except in the presence of a further stimulus from the same "dimension," to use Gibson's (1937) terminology. The mechanism of this lengthening, however, remains obscure.

### Experiment IV

To determine whether a movement after-effect which has come to an end during retinal anoxia induced after stimulation will reappear if the eye is released and the anoxic state terminated.

#### *Method and procedure*

Fourteen subjects with fairly long after-effects were again selected, the length of the after-effect seen with the unstimulated eye being the most relevant. Details of the apparatus were the same as in the previous experiment.

Length of after-effect was measured under three conditions:

1(*a*) in the opposite eye, after monocular stimulation;
1(*b*) when the after-effect ended the opposite eye was closed and the stimulated eye reopened; any further after-effect was noted;
2(*a*) after stimulation, pressure was exerted on the stimulated eye-ball, and the after-effect was reported with the opposite eye open;
2(*b*) when it ended, the subject covered the opposite eye, noted whether vision was completely lost in the stimulated eye, and then noted whether there was any further after-effect;

3(*a*) this was the same as Condition 2(*a*);

3(*b*) when the effect ended in the opposite eye, the subject released the stimulated anoxic eye and covered it, and noted whether any further after-effect occurred in the opposite eye.

The time normally elapsing between the application of pressure and loss of vision was determined for each subject.

*Results*

*Condition* 1*a*. An after-effect in the unstimulated eye was found in all subjects (they were selected on this basis).

*Condition* 1*b*. An after-effect was almost always seen with the stimulated eye after its ending was reported with the unstimulated eye. It was usually reported to be much more marked than it had been in the unstimulated eye.

*Condition* 2*a*. An after-effect was noted on all trials but one in the unstimulated eye while pressure was being exerted on the stimulated eye. After-effects in the opposite eye do not show a general tendency to be shorter than in Condition 1*a*, so it is not possible to draw any conclusions from this about the effects of anoxia.

*Condition* 2*b*. When the effect in the opposite eye had come to an end and pressure had been released, an after-effect was reported in the stimulated eye on both trials by 11 of the subjects.

An attempt was made to complete at least two trials in which the subject reported, on covering the opposite eye when the after-effect had ended, that he had in fact lost vision in the stimulated eye. This was sometimes difficult; the control observations on duration of pressure needed for loss of vision show this is frequently almost as long as the after-effect seen with the unstimulated eye and subjects sometimes reported that vision had only partially gone. However, two subjects reported spontaneously that they felt pressure on the stimulated eye was influencing the after-effect seen in the unstimulated eye.

The effects of anoxia are sometimes to be seen in the potentiating effect noticed in the previous experiment. For subjects 1, 3, 4, 5, 6, 7, 10 and 11 in Condition 2*b* the length of after-effect in the stimulated eye is greater than in Condition 1*b*. This suggests that retinal anoxia influenced the length of after-effect.

It also seems reasonable to infer that the after-effect must depend on retinal processes. If it were purely central the two eyes would presumably be equivalent after stimulation. One would not therefore expect that when the after-effect could no longer be seen with the unstimulated eye it could still be observed with the stimulated eye, but this was found to be so. This conclusion could also be based on the results of Condition 1*b*: the use of pressure adds little to the argument (except perhaps through its production of the "lengthening" effect, noted above).

*Conditions* 3*a and* 3*b*. Here a more convincing demonstration was attempted of the necessity of a functioning retina for perception of the after-effect of movement. The results show that in 7 of the 12 subjects taking part in Condition 3, *the after-effect in the unstimulated eye is dependent on the state of the stimulated eye*. The after-effect was seen to come to an end in the unstimulated eye during pressure on the stimulated one, and to reappear *in the same eye* when pressure was released. Incidentally, this result also supports the finding that it is very difficult to discriminate which eye is responding in certain conditions (Pickersgill, 1961).

The after-effect of movement seen with the unstimulated eye is, therefore, dependent on the state of the stimulated eye during the after-effect. The fact that this result is not capable of universal demonstration is not damaging to the conclusion;

the after-effect is always more difficult to see in the unstimulated eye and in a certain proportion of cases would have disappeared anyway before being interrupted by partial anoxia, so that it would not be expected to reappear afterwards. The important fact is that interruption and resumption of the after-effect in the conditions described can occur.

TABLE III

Lengths of after-effect in seconds on two trials measured under the following conditions:

(1*a*) in the opposite eye, after monocular stimulation;
(1*b*) in the stimulated eye, after any effect in the opposite eye had ended;
(2*a*) in the opposite eye, during pressure on the stimulated eye-ball;
(2*b*) as 1*b*;
(3*a*) as 2*a*;
(3*b*) in the opposite eye, after pressure on the stimulated eye-ball had been released.

| | Condition .. | | 1*a* | 1*b* | 2*a* | 2*b* | 3*a* | 3*b* | *Duration of pressure needed for loss of vision* |
|---|---|---|---|---|---|---|---|---|---|
| | N .. .. .. | | 14 | 14 | 14 | 14 | 12 | 12 | |
| | Eye .. .. | | U | S | U | S | U | U | |
| | P .. .. | | No | No | Yes | No | Yes | No | |
| | Measurement .. | | x | y | x | y | x | y | |
| | *Subject* | *Trial* | | | | | | | |
| Group 1 Reporting an after-effect in Condition 3*b*. | 3 | 1 | 11 | 9 | 11 | 15 | 14 | 8 | 6 |
| | | 2 | 10 | 9 | 12 | 16 | 15 | 11 | |
| | 5 | 1 | 6 | 8 | 8 | 21 | 3 | 10 | 8 |
| | | 2 | 6 | 7 | 3 | 17 | 3 | 6 | |
| | 7 | 1 | 8 | 10 | 7 | 11 | 6 | 6 | 3 |
| | | 2 | 6 | 10 | 8 | 11 | 5 | 6 | |
| | 9 | 1 | 12 | 10 | 5 | 7 | 6 | 7 | 8 |
| | | 2 | 5 | 11 | 4 | 11 | 6 | 7 | |
| | 11 | 1 | 13 | 15 | 24 | 23 | 20 | 20 | 15 |
| | | 2 | 17 | 13 | 23 | 21 | 18 | 15 | |
| | 12 | 1 | 5 | 8 | 7 | 0 | 4 | 13 | 6 |
| | | 2 | 3 | 10 | 6 | 0 | 6 | 8 | |
| | 14 | 1 | 5 | 8 | 12 | 0 | 6 | 12 | 8 |
| | | 2 | 9 | 8 | 6 | 1 | 10 | 12 | 5 |
| N=7 | | 1st Trial mean | 8·6 | 9·7 | 10·6 | 11·0 | 8·4 | 10·9 | |
| | | 2nd Trial mean | 8·0 | 9·7 | 8·9 | 11·0 | 9·0 | 9·3 | |
| | | 1st and 2nd Trials means | 8·3 | 9·7 | 9·7 | 11·0 | 8·7 | 10·1 | |
| Group 2 not reporting an after-effect in Condition 3*b*. | 1 | 1 | 9 | 9 | 8 | 10 | 7 | 0 | 10 |
| | | 2 | 7 | 9 | 0 | 18 | 7 | 0 | |
| | 2 | 1 | 3 | 17 | 8 | 19 | 6 | 0 | 6 |
| | | 2 | 2 | 13 | 6 | 13 | 5 | 0 | |
| | 4 | 1 | 3 | 15 | 6 | 16 | 9 | 0 | 5 |
| | | 2 | 2 | 10 | 8 | 16 | 6 | 0 | |
| | 8 | 1 | 5 | 9 | 14 | 7 | 5 | 0 | 6 |
| | | 2 | 6 | 8 | 8 | 10 | 4 | 1 | |
| | 10 | 1 | 15 | 15 | 10 | 19 | 12 | 1 | 4 |
| | | 2 | 15 | 14 | 11 | 16 | 9 | 0 | |
| N=5 | | 1st Trial mean | 7·0 | 13·0 | 9·2 | 14·2 | 7·8 | 0·2 | |
| | | 2nd Trial mean | 6·4 | 10·8 | 6·6 | 14·6 | 6·2 | 0·2 | |
| | | 1st and 2nd Trials means | 6·7 | 11·9 | 7·9 | 14·4 | 7·0 | 0·2 | |
| Group 3 not taking part in Condition 3*b*. | 6 | 1 | 10 | 0 | 8 | 5 | | | 5 |
| | | 2 | 13 | 0 | 7 | 10 | | | |
| | 13 | 1 | 10 | 8 | 11 | 0 | | | 5 |
| | | 2 | 9 | 12 | 8 | 8 | | | |
| N=2 | | 1st Trial mean | 10·0 | 4·0 | 9·5 | 2·5 | | | |
| | | 2nd Trial mean | 11·0 | 6·0 | 7·5 | 9·0 | | | |
| | | 1st and 2nd Trials means | 10·5 | 5·0 | 8·5 | 5·8 | | | |

*Key:*

U = the eye not stimulated by movement.
S = the eye stimulated by movement.
P = presence of pressure blindness in the stimulated eye while the measurement was being taken.
x = initial length of after-effect.
y = any further length of after-effect.

## Discussion

Amongst the early theories in which an attempt was made to explain the after-effect of movement, the main candidates were those stressing the importance of retinal factors and those associated with the ocular musculature and innervation. The question of cortical participation seems to have been mainly limited to consideration of whether or not attention influences the after-effect.

Some of the older theories necessarily involve localization of the after-effect, e.g. those which attribute it to fatigue in the ocular muscles, or to eye movements; the evidence for these is usually indirect, and it rarely sustains the interpretation put upon it. Such theories, in any case, have been applied only to the "waterfall" and not to the spiral after-effect.

Other theories, e.g. those of Wohlgemuth and Exner, suggest a mechanism which they consider would give an account of the physiological processes comprising the after-effect. A locus is suggested, but it is irrelevant to the explanatory value of the theory, and is largely guesswork.

More recent workers have indulged less in speculative theories of the earlier type, and have concentrated more on the investigation of specific variables. An exception to this is the more general explanation offered by Eysenck's theory of individual differences (1955*a*, *b*). It is assumed that when these differences manifest themselves in sensory phenomena the underlying causes of the differentiation are cerebral; the occurrence of "transfer" to the unstimulated eye is regarded as supporting evidence for this assumption.

If arguments from the occurrence of transfer are discarded as invalid grounds for inference, on the basis that an after-effect localized in the stimulated eye may well be reported as being in the non-stimulated eye, then there remains no previous experimental attempt to localize the origin of the after-effect.

The results of Experiment I show that while the report of an after-effect in the unstimulated eye does occur it is by no means universal. Holland (1957) states that the "transferred" length is about 70 per cent. of the normal length, and he implies that the relative contributions of retina and cortex can be estimated from this. The wide variation reported here alone renders this calculation suspect, as one would have to suppose the amount contributed by the retina varies from all to nothing in different subjects. The failure to discriminate which eye is the source of the after-effect may be due to the fact that the stimulus conveys a very reduced amount of information, since it is not affected in the same way as stimuli from material objects by e.g. eye movements, and closing the eyes.

The results of Experiments II and III suggest that the after-effect is dependent on the retina, since in certain conditions its presence or absence can be shown to depend on local oxygen deprivation.

These results, together with the finding that the after-effect of movement unlike that of light is not found after stimulation during pressure blindness, imply that the after-effect of movement may be determined by neural rather than photochemical changes; however, the fact that the after-effect can survive and is potentiated in length by moderate degrees of oxygen deprivation suggests that the explanation is a complex one.

One problem about ascribing the after-effect of movement to the retina is its duration. The upper limit has not been found, but after-effects lasting from 2 to 4 min. have been found after 10 min. stimulation. However, retinal processes with time courses from fractions of a millisec. up to over an hour can be found (Barlow *et al.*, 1957). Brindley comments "None fits the motion after-effect very well, but

I should not expect this as appropriate stimuli have never been tried" (Personal communication, July, 1959).[*]

It has been shown that the after-effect of movement can result from stimulation by stroboscopic pseudo-movement, and Exner showed that movement can be seen stroboscopically when the frames are presented to alternate eyes. This suggests the need, or at least the possibility, of cerebral co-ordination of successive stimuli for the perception of movement. There is, however, no inconsistency between this finding and the contention that a retinal contribution is essential during the perception of the *after-effect*. There is, in fact, considerable evidence that cerebral integration is more important than retinal for the perception of both apparent and real movement. It was shown by Smith (1948) that alternating stimulation of the two nasal retinae gave rise to the experience of apparent movement in 7 out of 10 subjects. The importance of cerebral factors has also been demonstrated in other studies: Gengerelli (1948) found the perception of phi was significantly better with stimulation of the same than of opposite hemispheres. This result was confirmed by Ammons and Weitz (1951); they also showed that phi was seen more often and appeared more stable with monocular than with binocular stimulation, indicating that retinal factors can play a part when they are allowed to do so. Descriptions of disturbances of motion perception in patients with brain lesions also point to the importance of cerebral factors (e.g. Ettlinger, 1955). At the same time it should be pointed out that an experiment by Smith (1940) provides evidence of a response to stimulation by movement which does not depend on the cortex. Guinea pigs respond to both rotating stripes and to steady stripes shown under intermittent illumination with pursuit movements of the head and eyes alternating with saccadic returns, and these responses persist after complete decortication.

Generalizations from guinea pigs to human beings must be made with caution, although Weiskrantz (1961, p. 57) has recently argued that "the effects of removal of the striate cortex appear to be surprisingly similar throughout the mammalian series." It may be that the perception of movement is comparatively independent of the cortex in human beings also; Riddoch (1917) has shown that perception of movement is the first function regained after occipital injuries causing the initial loss of all visual abilities: conceivably this is because it is organized at a subcortical level and is comparatively independent of the precise and complex organization of the cortex. It seems then that the most reasonable conclusion is that while the after-effect of movement is dependent on a functioning retina it is also dependent on the possiblity of normal movement perception, which to a much greater extent involves cerebral factors.

The writers would like to express their appreciation to Professor G. P. Meredith and Professor G. R. Hargreaves, in whose Departments these experiments were carried out, and also to the subjects, who took part in what were sometimes rather trying experiments. They are also grateful to Mr. Richard Gregory for helpful comment and discussion.

## REFERENCES

AMMONS, C. H., and WEITZ, J. (1951). Central and peripheral factors in the phi-phenomenon. *J. exp. Psychol.*, **42**, 327–32.

BARLOW, H. B., FITZHUGH, R., and KUFFLER, S. W. (1957). Changes of organisation in the receptive fields of the cat's retina during dark adaptation. *J. Physiol.*, **137**, 338–54.

---

[*] Barlow and Hill have recently reported the existence of ganglion cells in the rabbit's retina which are affected by prolonged movement stimulation; after 57 sec. stimulation their normal maintained discharge is suppressed for about 30 sec. See BARLOW, H. B., and HILL, R. M. (1963). Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. *Nature*, **200**, 1345–7.

Brindley, G. S. (1959). The discrimination of after-images. *J. Physiol.*, **147**, 194–203.

Cibis, P., and Nothdurft, H. (1948). Experimentelle Trennung eines zentralen und eines peripheren Anteils von unbunten Nachbildern. Lokalisation der Leitungsunterbrechung, die bei experimenteller Netzhautanäemie zu temorärer Amaurose führt. *Plüg. Arch. ges. Physiol.*, **250**, 501–20.

Craik, K. J. W. (1940). Origin of visual after-images. *Nature*, **145**, 512.

Craik, K. J. W., and Vernon, M. D. (1941). The nature of dark adaptation. *Brit. J. Psychol.*, **32**, 62–81.

Duke-Elder, W. S. (1926). The venous pressure of the eye and its relation to the intra-ocular pressure. *J. Physiol.*, **61**, 409–18.

Durup, G. (1928). Le problème des impressions de mouvement consecutives d'ordre visuel. *Année psychol.*, **29**, 1–56.

Ettlinger, E. G. (1955). *Defects of Visual Perception Associated with Cerebral Lesions in Man*. Ph.D. thesis, Univ. London.

Eysenck, H. J. (1955*a*). Cortical inhibition, figural after-effects and the theory of personality. *J. abnorm. soc. Psychol.*, **51**, 94–106.

Eysenck, H. J. (1955*b*). A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, **101**, 28–51.

Gates, L. W. (1934). The after-effect of visually observed movement. *Amer. J. Psychol.*, **46**, 34–46.

Gengerelli, J. A. (1948). Apparent movement in relation to homonymous and heteronymous stimulation of the cerebral hemispheres. *J. exp. Psychol.*, **38**, 592–9.

Gibson, J. J. (1937). Adaptation with negative after-effect. *Psychol. Rev.*, **44**, 222–44.

Granit, R. (1955). *Receptors and Sensory Perception*. Yale University Press, New Haven.

Helmholtz, H. von. (1924). In J. P. C. Southall (Ed.), *Treatise on Physiological Optics*. The Optical Society of America, U.S.A. 3 vols.

Holland, H. C. (1957). The Archimedes spiral. *Nature*, **179**, 432.

Holland, H. C. (1960). Measures of perceptual functions. In H. J. Eysenck (Ed.), *Experiments in Personality*, Vol. II. London.

Hunter, W. S. (1914). The after-effect of visual motion. *Psychol. Rev.*, **21**, 245–77.

Newton, I. (1831). Article 7. Experiments on ocular spectra produced by the action of the sun's light on the retina. *Edinb. J. Sci.*, **4**, 75–7.

Oswald, I. (1957). After-images from retina and brain. *Quart. J. exp. Psychol.*, **9**, 88–100.

Pickersgill, M. J. (1959). *The Determinants of some Aspects of Visual Movement, and their Relation to certain Personality Variables and to Brain Injury*. Ph.D. thesis, Univ. Leeds.

Pickersgill, M. J. (1961). On knowing with which eye one is seeing. *Quart J. exp. Psychol.*, **13**, 168–72.

Polyak, S. L. (1941). *The Retina*. Chicago.

Riddoch, G. (1917). Dissociation of visual perceptions due to occipital injuries, with especial reference to appreciation of movement. *Brain*, **40**, 15–57.

Smith, K. R. (1948). Visual apparent movement in the absence of neural interaction. *Amer. J. Psychol.*, **61**, 73–8.

Smith, K. U. (1940). The neural centres concerned in the mediation of apparent movement vision. *J. exp. Psychol.*, **26**, 443–66.

Weiskrantz, L. (1961). Encephalization and the scotoma. In W. H. Thorpe, and O. L. Zangwill (Eds.), *Current Problems in Animal Behaviour*. London.

Wohlgemuth, A. (1911). On the after-effect of seen movement. *Brit. J. Psychol. Monogr. Suppl.*, **1**.

*Manuscript received* 8*th September*, 1962.

# ERROR-CORRECTING PERCEPTUAL MECHANISMS

BY

D. P. ANDREWS

*From the Department of Communication, University of Keele*

The metric of the visual field is generally accurate but subject to changes dependent on stimulation. It is suggested that these changes indicate how the metric develops initially, and that the features of the visual signal which control the metric are the distributions of contours and of contour-curvatures over the field. These distributions are genetically prescribed for the organism; they correspond well with the distribution of these features in patterns derived optically from the environment; hence visual shapes and physical shapes are maintained in correspondence.

Further short discussions concern physiological hypotheses and the extension of the above to senses other than vision.

## INTRODUCTION

We take it for granted that a straight line *looks* straight, but this is really very surprising. The geometry of our private view of the world tallies rather well with that of the physical world revealed by measuring instruments; it also has the same number of dimensions. It follows that straight lines are possible in both; but that a straight line in the physical world is represented by a straight line in our internal visual world is very fortunate, and demands some fine engineering in the visual nervous system. We can also locate the position of an auditory click, touch finger to toe with the eyes shut, and perceive whether two receding lines are parallel. All these performances require very fine scaling in our internal representation of the world. How does the nervous system manage these things?

One possibility is that the system is just very accurately specified. This may be called the "telephone" theory; the whole wiring diagram of the exchange is specified genetically. Another view is that there are no pre-ordained connections; every cell is potentially in communication with every other, and the correct connections are set up laboriously by a learning process. Other views fall between the two, ascribing some connections to a processing of the input, and others to heredity; I wish to argue for a particular one of these, which I shall call the error-correcting theory.

For the greater part of this paper I shall discuss the perception of two-dimensional visual space, and finish with a more speculative discussion of steroscopic space and the metrics of other senses. Before this, I should like to introduce a few technical terms and say something about error-correcting devices.

A *metric* is the set of rules defining distances between points in any space. In three-dimensional Euclidean space, the metric is described by the rule that the distance between two points $(x_1, y_1, z_1)$ and $(x_2, y_2, z_2)$ is given by $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$. In other spaces Pythagoras' Theorem does not apply, for instance, in the two-dimensional space formed by the surface of a sphere. Its metric involves the radius. Some metrics are less easily described. Consider a figure drawn on a plane rubber sheet; if we now stretch some parts of the sheet, keeping it plane, the metric will be changed. To measure the parts of the figure after stretching, we should need to know the degree of stretch in two directions at each point; the whole table defines the new metric.

Surfaces which, if made of rubber, could be stretched without cutting to resemble one another, are said to be *Topologically Equivalent*, e.g. a sphere and a cube. If two metrics describe topologically equivalent spaces, we can devise a function which assigns each point in one space to a single corresponding point in the other. We shall call this a *transfer function*. For example, in the case of the rubber sheet the transfer function would describe where each point is to be found after the stretching. Given the metric of the initial state, the transfer function determines the metric after stretching.

An *error-correcting* device is one which accepts signals containing errors, and passes on a corrected signal. This it can do in three distinct ways:

(i) It "knows" the foibles of the transmitter, which for example always sends A for B and B for A; the inverse correction is made. This is analogous to the "telephone" theory.

(ii) The device has keys for making changes to signals received; a human operator records a large number of messages, and notes spelling errors and such. The errors are analysed, and the keys set up to correct them, when the operator is sure what the transmitter is doing wrong. This process could cope with complex conditional errors so long as they remained constant. This process is analogous to adjusting a metric by learning.

(iii) Messages are transmitted with *check symbols*. For instance 6-digit numbers could be sent with as many check digits. These represent the sums of various combinations of message digits, e.g. 1st check digit = 1st + 4th + 5th message digits.

Hamming (1950) has shown that this kind of process can be extended to reveal errors, both in position and size. A transmitter which sends check digits can therefore make errors in an irregular way and still be interpreted correctly. The correction can be made automatically and rapidly by any receiver which knows how the check digits work. This kind of error-corrector makes use of *redundancy* in the signal (as does (ii) in a rather different way); I shall try to show below that visual signals contain an analogous redundancy, which is used to correct the metric of subjective visual space continually.

### *Subjective metrics*

Perceptual theories describe the causal chain linking physical stimuli and perceptual experiences. The experimental techniques on which a theory rests may involve settings, judgements and statements by observers, but these need not form part of the theoretical model of perception; they are subsequent to the perceptual process and necessary only for transmitting the import of the subject's perceptions to other people. Some theorists insist that these actions be included in the model; this exercise can be made consistent, but I prefer the more direct approach. We think of visual perception in terms of the world we "see"; this is what we are trying to account for, regardless of philosophical creed. We may as well speak in such terms also.

However, it is necessary that any term we wish to apply to subjective space *can* be described in operational terms, lest it be dismissed as metaphysics. The redescription of phenomena such as illusions presents no problem; an operational definition of *subjective metric* is more difficult.

A subjective metric cannot always be established. Consider the scale of apparent loudness. So long as the greater of two physical noises always sounds the louder, the physical and subjective states are in topological correspondence; to determine the metric, however, we require an independent measure of subjective loudness. We can *adopt* a measure such as the j.n.d. or the apparent mid-point between two stimuli. We can then derive corresponding subjective metrics. But it is meaningless to ask whether Fechner or Stevens is *right*. To decide this, we still need that independent measure of subjective loudness. We may ask, which scale is most linear, most stable, or most consistent between persons; but these are criteria of convenience, not of truth.

In the case of visual space, a subjective metric can be established. We can compare any two lines and judge which is the longer. If one is adjustable, the lines can be made to appear equally long. The physical lengths and positions in space of the lines can then be measured. If we repeat this operation using various lines in various positions, we obtain a table of data which defines the metric of subjective space in terms of that of physical space. One figure is missing in the table—the size of the subjective units. This is strictly indefinable, but is in any case irrelevant to a perceptual theory; the *form* of transfer functions and other operations are the stuff of the theory. If we wished to evaluate the distortions of a camera lens, we might take a photograph of some simple geometrical objects. A print of *any size whatever* would serve to map the aberrations of the lens.

Visual space is not unique in this susceptibility. Subjective metrics of somesthesis, kinesthesis and sound-location can also be operationally defined. Such definition is only possible when perceptual and physical objects can move in their respective spaces without losing their identities. We can equate two lines wherever they are in the visual field, and then measure them. Measurement is possible only because we know that physical objects move without changing shape; this measuring can be reduced to detecting the coincidence of points; this performance is unaffected by any distortions of perception.

In the scale of apparent loudness, these features are absent. We do not know how large an interval at the quiet end of the scale is equivalent to a given interval at the loud end, because there is no corresponding physical object known to be unchanging which can be moved from one position to the other. Even if a person tries to make such judgements, we do not know what his criteria are, nor what they ought to be.

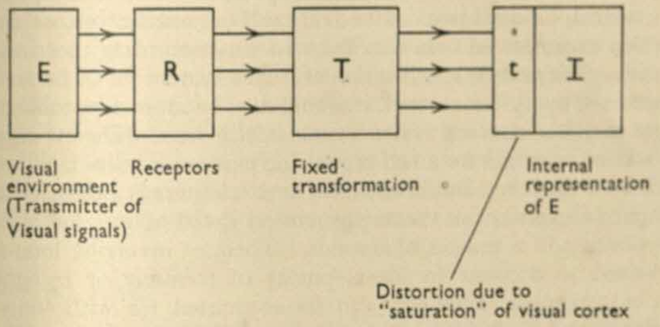
The scale of pitch is perhaps an intermediate case; we can use the octave as a movable object, though the process does have something of the arbitrariness of using the j.n.d. for the same purpose in the scale of apparent loudness. Pitch owes its measurability to its relation with time interval, which is common to both physical and subjective worlds; the quality of intervals such as octaves is both subjectively distinctive and related to time interval in a simple way.

## *Some theoretical models of visual space perception*

Our private visual space (henceforth referred to as (I)) has some remarkable peculiarities. We experience optical illusions, figural after-effects, and irregularities of metric, such as the horizontal-vertical illusion and Luneberg's (1947) finding that visual space is hyperbolic.

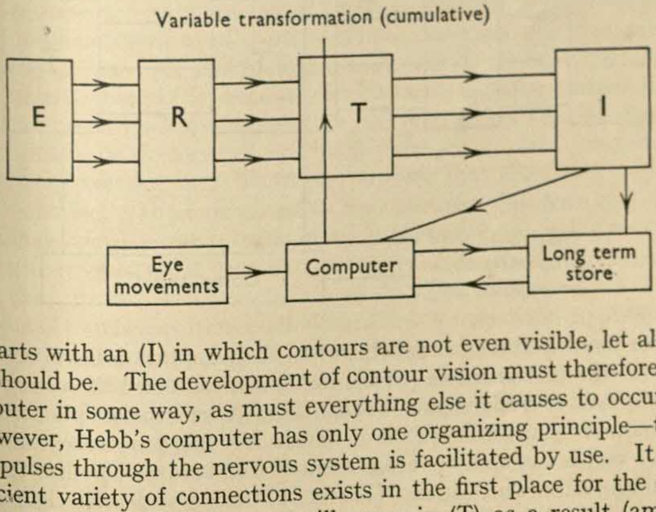
Theories of space-perception are largely theories to explain such peculiarities. One way of explaining these phenomena is to consider the visual system fixed, but subject to distortion in special circumstances. A good example of this approach is Köhler and Wallach's (1944) theory of figural after-effects (see Fig. 1). (E) is the visual environment, a pattern of light. It might be regarded as the set of all possible photographs. (R) is an ideal receiver (the eye) which takes samples from available light patterns and codes them as neural activity. Other operations concerning space perception, which may in fact occur in the eye, are referred to (T) and (t). (T) is the fixed transfer function of the visual system. Its connections ensure a correct transfer of metric between (E) and (I). (t) represents "saturation" in the visual cortex, which distorts the metric of (I) temporarily. This occurs in the neighbourhood of stimulus contours. After the saturation has worn off, the overall transfer function (T + t) goes back to what it should be, (T). The transformation (t) is a consequence of the shortcomings of the visual cortex in this theory, and is unhelpful to the organism.

FIGURE 1


E
R
T
t
I
Visual environment (Transmitter of Visual signals)
Receptors
Fixed transformation
Internal representation of E
Distortion due to "saturation" of visual cortex

At the other extreme is Hebb's (1949) theory of visual space perception (see Fig. 2). (R) serves the same function in this diagram; all processing other than coding is referred to (T) or (C). In this theory the transfer function (T) is supposed random at birth; a signal at (R) causes activity over a considerable region of (I), but with some topological correspondence preserved. (C) is a computer which controls (T). (C) utilizes all the input to the organism throughout its history. It has the job of establishing exact topological correspondence between (R) and (I), and developing a metric for (I). This process is supposed to take months.

FIGURE 2

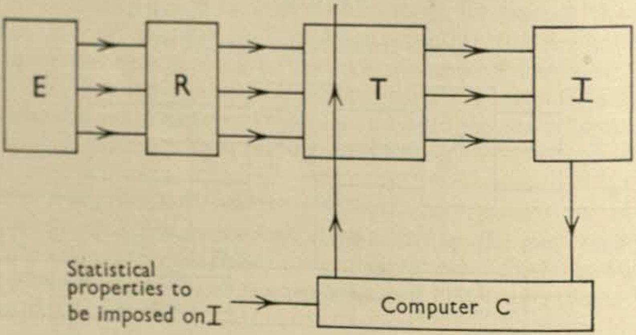


Hebb starts with an (I) in which contours are not even visible, let alone straight when they should be. The development of contour vision must therefore be inherent in the computer in some way, as must everything else it causes to occur or develop in (I). However, Hebb's computer has only one organizing principle—that a given path for impulses through the nervous system is facilitated by use. It is supposed that a sufficient variety of connections exists in the first place for the correct ones to develop. It is clear that changes will occur in (T) as a result (amongst other changes), but we require a special kind of change—one which makes (I) become progressively more like (E). This feature Hebb fails to demonstrate convincingly. His computer cannot distinguish good and bad changes, but merely exaggerates progressively such changes as occur. The system requires a reference criterion, and would be unstable without one. Hebb offers eye-movements in this role; he suggests that the eyes tend to move so that prominent features such as corners fall in the centre of vision. The perception of contours, *inter alia*, is supposed to develop in consequence; but I find this part of his argument very unconvincing.

This kind of computer could be made more plausible if more computer characteristics were specified. No system can be truly self-organizing, but we could devise a computer which organizes its own detailed structure according to principles built-in somewhere else. The genetic specification of such a system would be far shorter than one which specified every connection; it would also be more adaptable, and it seems likely that much of the nervous system works in this way. The eye-movement idea is the right sort of principle for a self-organizing computer, as is facilitation. These are just too few for the job. But in any case, such a system is insufficiently dependent on recent input to explain the visual phenomena listed above; for instance, figural after-effects develop in a matter of seconds. Theories involving long-term storage are better suited to account for development of meaning or recognition in (I). Pecularities of perceptual geometry can be accounted for with only short-term storage available, and used in an automatic way.

There is a middle course between theories of learned and genetically prescribed space perception. It is arguable that the topology of visual space is genetically determined, and the metric due to some form of learning. There is strong physiological evidence that an important degree of topological correspondence is preserved throughout the visual nervous system. Also, Lashley (1942) has shown that rats at least can perform complex perceptual tasks with only a few minutes available for learning. Let us presume therefore that (E), (R), and (I) are in topological correspondence to start with, and attend only to the metric.

FIGURE 3



Helson has developed a concept which he calls Adaptation Level (1948). Colour adaptation was the first example (1938); he later generalized the concept to include other psychophysical judgements (1947), and even social phenomena (1940). Figure 3 illustrates the idea. (T) is a device for assigning categories to stimuli; for example a number of weights are presented to a subject who has to assign them to one of 11 subjective categories. (I) is a subjective scale of weight divided into 11 categories. Initially, the boundaries are undefined. After some experience of the range of values to be expected, the subject is able to classify the weights consistently. Helson found that the *middle* category was given to a weighted mean function of all the stimuli presented. This criterion is represented as an independent statistical input to (C); (C) also computes this weighted mean function.

Helson found that colours are averaged in a similar way. (R) receives colour stimuli, and (T) transforms all colours according to instructions from (C). (C) computes the average colour seen throughout the visual field in (I) at any one time, and modifies (T) to make this average hue more nearly colourless. (All other colours are modified to a corresponding extent.) The change is very rapid. The statistical criterion here is that average hue tends to appear colourless.

The Adaptation Level concept can be extended in one important respect. There are no absolute scales of heaviness or hue, so changes in these subjective scales are neither correct nor incorrect. The reference criterion in these two examples cannot represent the "true" state of affairs in the environment. However, a metric presenting optical illusions is clearly in error compared with one that does not. A similar device might organize a *correct* metric of visual space if we can find a "true" criterion. The uniform distribution of some feature of the visual signal might serve this purpose.

*The error-correcting model*

The principle can be illustrated by considering the metric for apparent movement in (I). The waterfall illusion can be explained in this way (see Fig. 3).

(T) transforms the appearance of motion in (I). It has a unique but initially arbitrary transfer function, so that at birth a stationary scene in (E) might be represented at (I) having illusory movements distributed randomly over it. The transfer function of (T) is continually computed by (C) which has two inputs:

(1) Data derived from (I) concerning ds/dt for contours. ((I) is assumed complete except for representation of motion).

(2) A statistical criterion determining how motions are going to be distributed at each point in the field of (I). (This criterion is genetically prescribed).

The criterion is that the average movement over a period for every element in (I) is to be represented as zero, i.e. *(T) will be modified from its initial arbitrary characteristic until such time as all parts of (I) are exhibiting zero average motion in all directions.* Thenceforth, (C) will maintain this state of affairs. Now, this happens also to be generally true of (E), because objects move as often one way as the other; similarly with eye-movements. Therefore the imposition of this statistic on (I) would not normally be detected so long as (C) computes over a long enough period for movements to average out. However, suppose that there is a waterfall in (E), and that it is fixated. The average motion in the corresponding part of (I) will be *continually changed by (C) until it is stationary.* (Actually, this point is never reached, but a considerable reduction of seen movement occurs.) Now, the eyes are moved to a stationary background; the characteristic of (T) in the affected region, is subtracting a certain value of movement from all motions, zero or otherwise, so the corresponding part of the background appears to move upwards. Immediately (C) begins to recompute the transfer function, and the illsory movement gradually diminishes.

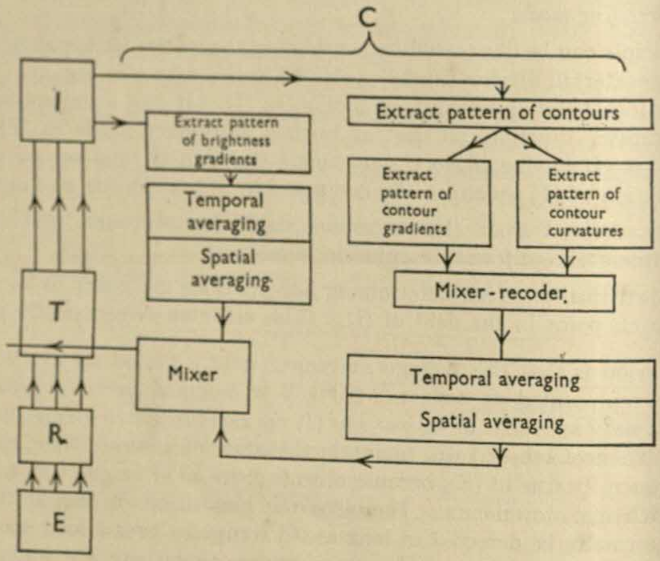
This device is error-correcting only in that *stationary* is substituted for mean ds/dt. It is not necessary for this device to work that ds/dt values reaching (I) should be correct—only that their errors should be consistent over a period. Since visual transmission lines are long, and speeds of nerve conduction vary, this device frees the nervous system from requirements that might be difficult to engineer.

The averaging process mentioned above has not been specified, either in respect of the time interval or the lateral spread over which it takes place. I shall describe the detailed operation of (C) in the next example.

For most of the time the metric of visual space corresponds rather well with that of (E). Kohler (1951) has shown some startling visual phenomena of adaptation. For instance, if we assume distorting spectacles, (R) gives a distorted output at first reflected in (I). This distortion gradually goes, and despite the distorting lenses, we see straight lines as straight once more. It follows that the correspondence of (E) and (I) is not the result of a constant transfer function, and this applied equally to the case before the distortion was introduced. How, therefore, are straight lines seen as straight in the first place?

Consider Figure 3 once more, this time in the guise of an organizer of spatial metric. (I) has an initially arbitrary metric; topological equivalence is preserved, however. (C) *is to change the metric in* (*I*) *until the contours there are on average straight in every part of the field.* It operates as follows. The right-hand loop of Figure 4 illustrates the detailed operation. (C) extracts from (I) a pattern of contours (loci of high brightness gradient dL/ds). From this pattern it extracts a pattern of contour curvatures and a pattern of contour gradients. Gibson (1950) has already suggested that the visual system extracts this kind of data, for the perception of depth. The patterns of contour curvature and gradient are then transformed into two patterns representing curvature with respect to two directions in the field, say vertical and horizontal. (The directions to be used are arbitrary, and may vary from point to point in the field.) (*C*) *can now associate any point in* (*I*) *with two values which represent its contour curvatures.*

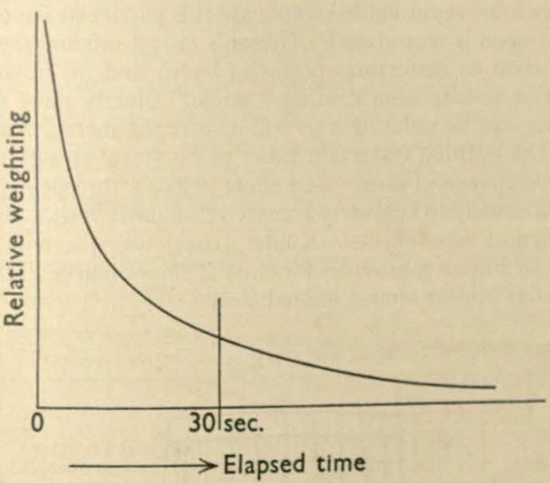
FIGURE 4


C
I
Extract pattern of contours
Extract pattern of brightness gradients
Extract pattern of contour gradients
Extract pattern of contour curvatures
Temporal averaging
Spatial averaging
Mixer - recoder
T
Mixer
Temporal averaging
Spatial averaging
R
E

These curvature values are averaged over a period, with a time-weighting function as in Figure 5. Recent data are heavily weighted. These weighted mean values are then averaged laterally, to compute curvature values for each point dependent on neighbouring points also. The closer points contribute most; the weighting function may again be represented by Figure 5 with distance in place of time. (C) now has two values for any point in (I), which represent the curvatures in the neighbourhood centred there, with a strong bias to proximity; past and present stimuli contribute, with a heavy bias to recency.

These two values control the transfer function, and hence determine the spatial metric of (I). The criterion for the change is that it continues until the weighted mean curvature values are zero. The result of all this is that the metric available at (R) is never used; instead is substituted a metric depending on the statistical distribution of contour curvatures. Fortunately, we gain by the substitution; (I) generally comes out to be more like (E) than (R's) output, as is shown by the experiment with distorting spectacles. *The statistic imposed on (I) is true also of (E) on average*; a curvature is as likely to be concave as convex, so the mean value is zero over a period involving normal eye movements.

FIGURE 5



It will not have escaped any geometers that the device has not yet defined subjective space uniquely. Any conical projection preserves straightness, so (I) might exhibit any of an infinity of conical projections of objects in (E). We require a second stage; without it we should be unable to compare distances in our subjective visual space. The second part transforms (I) by means of (T) also, so that *on average, contours are equally spaced throughout the field.*

From (I) is derived a pattern of brightness gradient at each point, with no regard to sign. *Now, for a field of uniform texture in random motion in (E) this function would be evenly distributed on average.* Here we have a statistic we can impose on (I). This function is averaged spatially and temporally as for the curvatures, and (T) transforms the metric of (I) so that the second spatial differential of the average values is zero. The metric of (I) is now specified redundantly by the composite device, and any incompatibility must be prevented. Let the final transfer function be that combination of the two described above, which provides the maximal correlation with each. This need for compromise might suggest that these two mechanisms are needlessly powerful; the choice of contour curvature and contour spacing was not necessary, and other derivatives of the signal might prove better.

I selected these two only because they give a simple account of illusions, figural after-effects and other visual phenomena.

This complete mechanism has two notable properties:

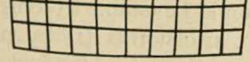
(i) Curves will tend to get straightened out in (I), whenever the perceptual diet from (E) is biased with curves all one way for an appreciable period.

(ii) Regions in (E) of high brightness gradient will be given expanded representation in (I), i.e. any region where a contour falls continuously will become stretched at right angles to the contour.

It is easy to show that such a device exhibits figural after-effects. The "repulsion" of visual objects by a heavy contour follows from (ii). A weak contour suffers more attenuation by the spatial averaging, and also by imperfect fixation. The "distance paradox" follows, because the integral of scale-change × distance (i.e. apparent displacement) will be maximal some distance from the contour causing the disturbance. There should be no net after-effect with coincident (I) and (T) figures, though eye-movements of any kind will produce some. The opposing sides of closed figures will produce some repulsion to complicate this particular issue. The tendency for small angles to open is reproduced. Gibson's (1953) original after-effect is given by (i), and adaptation to distorting spectacles by (i) and (ii) in combination. An interesting distortion results from fixating a circle. Clearly there is a limit to how much the curvature can be reduced; this will distort the metric in a way not underwritten by (ii). The optimal distortion takes place—local straightening of parts of the circle with little change of area. The effect is like a threepenny bit.

The view seems already to be widely accepted that most optical illusions are simply special cases of figural after-effects. Köhler (1944) was the first to suggest this. The characteristic of Figure 5 provides for this; it shows marked bias to very recent data, so that illusions appear almost immediately.

FIGURE 6



The second condition imposed on (I) has some interesting side effects. It results in contours being evenly distributed over (I). Now this is only approximately true of (E) at the best of times. Besides abnormal biases introduced by fixation, biases are present in a normal input of varying selections from (E). In the upper part of the field, contours tend to be widely spaced due to the absence of objects. In the lower part of the field they are rather more numerous. In the central band of the field, they are more numerous still because distant objects present closely spaced contours, and distant objects are more likely to fall in the centre of the field.

However, our device can know nothing of all this, and will not make appropriate compensations; so the upper part of the field, in (I) is contracted, the middle part expanded, and the lower part slightly contracted, as compared with a photograph of (E). These contractions and expansions take place both upwards and sideways; this will give rise to spurious curvatures of contours, so the curvature sensitive device will reduce the effect. The final result is illustrated in Figure 6, which shows how a uniform grid would appear after a normal perceptual diet. (The effect is exaggerated.)

An object high in the field looks too small. The width of a square looks less than its height—the horizontal-vertical illusion. The illusion is reversed in the upper part of the field. All these phenomena occur; they were known to the ancient Greeks and their architecture incorporated the inverse effects to make the lines look right. Finally, the whole metric is distorted in the way Luneberg found. His colleagues (Hardy *et al.*, 1953) have established since his untimely death, that visual space is not precisely hyperbolic; if the present explanation is correct, this is hardly surprising. (The third dimension has not been discussed. If we assume it is derived from the differences between two two-dimensional metrics distorted as above, visual space will be approximately hyperbolic.)

## DISCUSSION

This theory is a functional theory; it describes a series of quasi-mathematical operations. These operations could be incorporated in numerous hypothetical physiological mechanisms. I do not feel qualified to press any particular ones, but I should like to throw out a few suggestions. Operations that are complex to describe in terms of integration, differentiation, etc., may be embodied in simple and attractive physiological devices.

As an engineering problem, the transmission of the metric through the visual nervous system is peculiarly difficult; it is likely to suffer distortion at the retina, where some cells are larger than others; in the optic nerve, where the periphery claims far fewer fibres than the fovea; and at relay stations further in, where varying connectivity occurs. There are two types of solution to the problem. (i) To connect each nerve to the *right* place in whatever gives rise to the metric of (I). This would require the visual system to be made with extraordinary precision, and is in any case inconsistent with adaptation of the metric. (ii) To code the metric at the outset in a way not subject to the twists and stretchings of the visual nervous system. Such a code is fortunately provided by the environment in the redundancy described above, so the visual system is spared the job of coding. It is worth reflecting, how could the job be done otherwise? A receptor cell is well adapted to signalling "I am being stimulated"; but how could it say "I have six neighbours, each 3 microns from me and from each other." This is the sort of information that would have to be coded to transmit the metric itself, unless we resort to a "telephone" theory.

Lateral inhibition could conceivably be cast in this role; time interval might represent distance for the purpose. However, adaptation to distortion would be hard to account for. Nevertheless, lateral inhibition is probably part of the story. Barlow (1961) has shown that its effect is to reduce the whole visual signal to a pattern of contours, which is exactly what is required for the first stage of the computer (see Fig. 4). Lateral inhibition involves summation (Barlow, 1953; Hartline and Ratcliff, 1958). The latter by itself is required in later parts of the computer. Lettvin *et al.* (1959) have shown that there is a specific response to contour curvature in the frog's optic nerve. Hubel and Wiesel (1959) have demonstrated specific responses to contour-gradient in the striate cortex of the cat. All this suggests that much of the computer and transformation machinery may be housed in the retina, for the periphery at least; processing of foveal signals appears to occur at a later stage in the human visual system.

If the visual system is unable to code such information then can it decode it? It is not necessary that the description of the metric be decoded into actual physical distance; any sufficient description will do, in whatever dimensions. We already have such a description in the average curvature and contour-spacing values which control (T). It is as though we were to describe one possible state of a rubber sheet by giving the tension in two directions at every point.

(I) and (T) may actually be layers of cortical cells in the visual projection area. The activity of the T-layers represents the local expansion in the spatial field, and the I-layers display the signals coming in. Two of these in conjunction might give rise to stereopsis, and others represent colour and movement. This arrangement would have the advantages (i), there are no transmission lines between (I) and (T); (ii), the fovea can be represented by a disproportionately large area without affecting the metric in any way.

I should like to reconsider for a moment the error-correcting decoder of digits introduced earlier. The principle of operation is that the signal contains check digits. The decoder knows what the intended relations are between the check digits and the message digits. The signal is therefore redundant, and it is this which makes error-correction possible. The visual signal at (E) contains redundancy also, in that the incidence of contours and their curvatures, over a period of time, is largely independent of the specific data involved.

It has been shown how the metrics of the two-dimensional visual field and visual motion perception might be derived from redundancy of the stimulus pattern. This concept can be extended to other metrics. Wherever there is a projection area in the cortex, there is a metric to be established, either by preordained connections, by learning, or by autonomous response to stimulus patterns. Kinesthesis, somesthesis, pitch perception, sound location, and stereoscopic vision, all have accurate metrics. Let us consider these briefly in turn.

Let us suppose that initially there is a body image and kinesthetic representation, however distorted. The actual relative motion ds/dt between any two parts of the body is zero on average over a period. This is true for each possible relative position of the two parts. Any cyclic movement has no net displacement; displacement is the integral of ds/dt. Let us suppose the nervous system brings these things about also in the subjective representation of body movements. Metrics for somesthesis and kinesthesis will be continually modified. The averaging will make kinesthetic after-effects possible, if we assume a non-linear time-weighting function as in Figure 5. These have been observed. The metrics would be subject to distortion in states of reduced stimulation following amputation.

Licklider (1959) has shown how musical intervals can be established by autocorrelation; perhaps harmonic ratios serve as the pattern for organizing the metric of pitch. Harmonics are commonly present together. This would be consistent with the special quality of octaves, the qualities of musical intervals and the absence of these effects at very high frequencies where harmonics are beyond the audible range.

Sound location presents a similar problem to that of the visual field. The distribution of transients around a person will be approximately uniform over a sufficiently long period; if this were brought about for the subjective metric, it would be fairly accurate; there should be an interesting after-effect if the metric is established in this way. This metric is particularly in need of explanation, as the effective time interval for locating a click is small compared with the time required for transmission; any fixed metric device would have to be incredibly precise, whereas an adjusting-metric device does not even require equal rates of nerve-conduction in the whole auditory channel.

Finally, the metric of stereoscopic depth may be organized so that texture is on average uniform throughout the third demension. Gibson (1950) has presented strong evidence that texture gradient is extracted from the visual signal, and used to perceive objects at the correct distance; the same data may very well be fed to a device which organizes the metric itself.

These latter remarks are highly speculative, though I hope stimulating. In particular, the kinesthetic metric is too complex a problem to be stated or solved in one paragraph. My intent is only to show that this approach is available with the metrics of other senses. Whether or not all senses operate in this way, there seems to be a strong case that the visual metric is derived from the redundancy of stimulus patterns.

My thanks are due to Professor D. M. MacKay, Mr. R. J. Paul, and Dr. H. Liebeck for patient study of the script and many helpful comments.

## References

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *J. Physiol.*, **119**, 69–88.

Barlow, H. B. (1961). In W. A. Rosenblith (Ed.), *Sensory Communication.* New York. 782–86.

Gibson, J. J. (1953). Adaptation, after-effect and contrast in the perception of curved lines. *J. exp. Psychol.*, **16**, 1–31.

Gibson, J. J. (1950). *The Perception of the Visual World.* Boston.

Hamming, R. W. (1950). Error-detecting and error-correcting codes. *Bell System tech. J.*, **29**, 147–50.

Hardy, L. H., Rand, G. M., Rittler, M. C., Blank, A. A., and Boeder, P. (1953). *The Geometry of Binocular Space Perception.* New York.

Hartline, H. K., and Ratcliff, F. (1958). Spatial summation of inhibitory influences in the eye of Limulus, and the mutual interaction of receptor units. *J. gen. Physiol.*, **41**, 1049–66.

Hebb, D. O. (1949). *The Organization of Behaviour. A Neuropsychological Theory.* New York.

Helson, H. (1938). Fundamental problems in colour vision. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *J. exp. Psychol.*, **23**, 439–76.

Helson, H. (1940). Fundamental problems in colour vision. II. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *J. exp. Psychol.*, **26**, 1–27.

Helson, H. (1947). Adaptation level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, **60**, 1–29.

Helson, H. (1948). Adaptation level as a basis for quantitative theory of frames of reference. *Psychol. Rev.*, **55**, 297–313.

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.*, **148**, 574.

Kohler, I. (1951). *Über Aufbau und Wandlungen der Wahrnehmungswelt.* Vienna.

Köhler, W., and Wallach, H. (1944). Figural after-effects. *Proc. Amer. Phil. Soc.*, **88**, 269–357.

Lashley, K. S. (1942). The problem of cerebral organization in vision. *Biol. Symp.*, **7**, 301–22.

Lettvin, J. T., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proc. Inst. Radio Engs.*, **47**, 1940–51.

Licklider, J. C. R. (1959). Three auditory theories. In Koch, S. (Ed.), *Psychology. A study of a Science.* Vol. I., 41–139. New York.

Luneberg, R. K. (1947). *Mathematical Analysis of Binocular Vision.* Princeton.

*Manuscript received 24th September, 1962.*

# IS THE WEBER FRACTION A FUNCTION OF PHYSICAL OR PERCEIVED INPUT?

BY

HELEN E. ROSS and R. L. GREGORY

*From the Psychological Laboratory, University of Cambridge*

It is well known that a given physical input (e.g. intensity of light or sound, length or weight of an object) does not always give rise to the same sensation. For example, arrow heads affect the perceived length of lines (Muller-Lyer illusion) and size affects the apparent weight of an object (size-weight illusion). It is generally assumed that the differential threshold is a simple function of the physical intensity of the stimulus. We may however ask whether the differential threshold is affected by illusions.

To try to answer this question we estimate differential thresholds under a pair of conditions in which the relevant input is the same but appears different. Using the size-weight illusion, we have found that the differential threshold for weight is greater for a set of small weights than for a set of larger and apparently lighter weights.

## INTRODUCTION

A basic and much studied aspect of sensory processes is the ability of the nervous system to distinguish small differences in the intensity of physical inputs. The Weber fraction $\triangle I/I = C$ is found to hold fairly accurately over a medium range of intensities for most modalities (Woodworth, 1938). Sensory thresholds are now regarded as the distinguishing of signals from noise according to some significance criterion (Tanner and Swets, 1954; Gregory and Cane, 1955; Gregory, 1956; Barlow, 1956). On this view we may think of the Weber Constant in terms of detector efficiency. There are many known factors which affect human detector efficiency—such as fatigue, vigilance and sensory adaptation—which are due to changes in the organism.

Estimates of differential thresholds are generally made in terms of the physical input, using physical scales of weight, loudness or photic intensity. This raises problems concerning the relation between physical and psychological measuring scales into which we do not propose to enter here; but the well known fact that a given physical intensity may give rise to different sensations under various conditions raises a question which seems to demand an experimental treatment. Discrimination is affected by the psychophysical method used (Guilford, 1936), the instructions given to the subject (Fernberger, 1931) and the subjective probability of each decision (Howarth and Bulmer, 1956); but we may also ask whether it is affected by illusions. It is well known that a given intensity of light may produce more or less sensation of brightness as a function of contrast and that a weight may feel light or heavy depending on its size (size-weight illusion). We thus require the idea of *apparent* intensity as well as physical intensity. We say that two weights are apparently equal when they are judged the same weight by a subject. It is found that if one of two physically equal weights is larger in size than the other it will be judged lighter, so in this case physical equality (balancing on weighing scales) is not the same as apparent equality. This difference between physical and apparent equality is what we mean here by an illusion.

The fact that physical and apparent intensities may be systematically different under certain conditions makes it possible to ask whether sensory discrimination is a simple function of physical intensity, or whether it may be affected by changes in the apparent intensity produced by conditions giving illusions.

The size-weight illusion was chosen to investigate this question as it is large and requires the simplest apparatus. It is only necessary to obtain differential thresholds for weights of different sizes but the same physical weights to establish whether the threshold is a function of the physical or the apparent weight. This question has been investigated by Seashore (1896) using a similar method to ours, but we cannot place much confidence in his results owing to his method for determining the thresholds, and we find his argument self-contradictory. He suggests that Weber's Law is better fitted by the apparent than the physical weight, and that the Weber fraction should be $\Delta$ I/apparent I = C. Surprisingly, he concludes that "all over-estimation (of weight) lowers the threshold and all underestimation raises it." This contradicts his own formulation. Seashore presents two measures of thresholds. Using the limiting method, he accepted the first of three consecutive correctly judged increments; he also gives proportions of incorrect to total number of judgements, though his procedure here is not entirely clear. He gives no indication of the statistical significance of his results, and his two sets of measures go in opposite directions.

We originally intended to perform only one experiment, using the frequency method and presenting the standard before the comparison weight. A second experiment was done with the order of presentation inverted in order to discover what happens to the constant error in these conditions, and finally a third experiment was undertaken lifting the weights simultaneously in order to eliminate constant error. In order to reduce to a minimum the number of readings required for an adequate significance we had to find a method giving a small variance. It turned out that methods 1 and 3 were satisfactory, while the method used for Experiment 2 was unsatisfactory since with this method the variance was high.

The results of all three experiments will be given; it should be understood that they were not designed to give answers to different questions but are attempts to find the best method to answer the original question.

## METHOD

Two sets of weights were used: one set 3 in. × 3 in. (Nescafé tins), and the other set 1¼ in. × 1¾ in. (Kodak film tins). All the tins were painted black. Each set was composed of 11 tins which were weighted with lead shot, wrapped in cotton wool, to weigh from 175–225 gm., with 5-gm. intervals between them. The 200-gm. tin was used as the standard for the tins of the same size.

It was essential for the logic of the experiment that there should be no difference between the series of weights, except for size, which could favour discrimination for either series. It was, therefore, important to standardize the finger-span required to lift the weights, although they were of different diameter in the two series. This was done by fitting the lids of all the tins with identical wire loops, which were grasped by the thumb and forefinger.

Forty-seven subjects were used in the experiments, five of whom served in more than one experiment. They were of both sexes, aged between 17–30 years, and were mainly undergraduates, research students and college entrance candidates.

The subject was seated at a table with one of the standards in front of him. A screen hid all the weights except the pair being compared. The weights were placed in turn beside the standard in a predetermined random order. Each weight appeared 10 times, providing 100 comparisons for each set of weights. The subject was asked to compare each weight with the standard for that series, and judge whether it was heavier or lighter than the standard. If he was not sure, he was asked to guess, thus eliminating the variation in threshold caused by different interpretations of the "equal" category.

Three experiments were undertaken. The conditions in each experiment were the same except for the order in which the standard and comparison weights were lifted. In all cases the subjects were required to hold the weights by the wire loops, using the thumb and forefinger, and to lift and lower them smartly and as consistently as possible. The preferred hand was used throughout, except in Experiment 3, when a weight was held in each hand. All subjects were tested on both sets of weights; half being tested first on the large weights and half first on the small weights, although it turned out that order was unimportant.

### EXPERIMENT 1 (Standard first, comparison second)

Using the procedure described above, the subjects lifted the standard first and the comparison weight second. Sixteen subjects were used.

### *Results*

In treating the results non-parametric statistics were used, to avoid assumptions of normality. Unless otherwise stated, the significance of the difference found between the two sets of weights was estimated by the Wilcoxon matched-pairs signed-ranks test (Siegal, 1956, p. 75). The detector efficiency is given by the slope of the probability curve for the response "heavier" or "lighter"; it does not matter which. The standard deviation (SD) (i.e. the slope of the probability curve) and the point of subjective equality (PSE) were calculated for the large and small weights using the method of averaged z scores as described by Woodworth and Schlosberg (1955, p. 205). The use of this method results in a slight overestimation of the SD for displaced PSEs. However, the overestimation is small compared to the differences in SD obtained, and we accept the method as sufficiently accurate.

For the large weights the median SD was 11·9 gm., and for the small weights 13·8 gm. The probability of this difference being due to chance was 0·03 (one tail).

It was also found that the median PSE was different for the two series. For the large weights the PSE was 195·2 gm., and for the small weights it was 192·1 gm. This difference in "time error" is significant at the 0·02 level (two tails).

TABLE I

EXPERIMENT 1: STANDARD FIRST, COMPARISON SECOND

| *Subject* | *PSE* | | *SD* | |
|---|---|---|---|---|
| | *Large* | *Small* | *Large* | *Small* |
| 1 | 192·1 | 187·6 | 12·7 | 11·8 |
| 2 | 201·8 | 196·3 | 10·2 | 8·0 |
| 3 | 193·3 | 192·8 | 9·4 | 10·5 |
| 4 | 198·0 | 190·2 | 15·1 | 19·1 |
| 5 | 190·8 | 186·0 | 13·7 | 13·8 |
| 6 | 207·1 | 199·8 | 10·7 | 14·6 |
| 7 | 195·0 | 190·7 | 10·1 | 12·6 |
| 8 | 192·6 | 194·3 | 12·6 | 13·0 |
| 9 | 196·0 | 199·0 | 12·2 | 16·2 |
| 10 | 202·3 | 187·6 | 11·9 | 13·8 |
| 11 | 194·5 | 199·5 | 14·2 | 12·0 |
| 12 | 196·8 | 193·8 | 10·3 | 14·0 |
| 13 | 190·4 | 191·5 | 10·6 | 22·2 |
| 14 | 195·3 | 190·4 | 9·4 | 14·9 |
| 15 | 190·2 | 190·0 | 15·3 | 13·8 |
| 16 | 203·0 | 199·3 | 11·9 | 10·7 |
| Median | 195·2 | 192·1 | 11·9 | 13·8 |

Significance of difference between PSE, $p = 0·02$ (Wilcoxon).
Significance of difference between SD, $p = 0·03$ (Wilcoxon).

It seems that this experiment provides evidence that the smaller (apparently heavier) weights are more difficult to distinguish from their standard than are the larger (apparently lighter) weights from theirs. This is a positive result, and suggests that detector efficiency is related to apparent weight. Since this is a somewhat startling conclusion it seemed necessary to obtain additional data.

Experiment 2 was undertaken with the comparison weight presented before the standard in order to obtain more data on the thresholds but with the time error reversed.

### EXPERIMENT 2 (Comparison first, standard second)

The procedure was exactly the same as in Experiment 1 except that the comparison weights were presented in each case *before* the standard. Sixteen subjects were used.

*Results*

The subjects found this condition very difficult: they were instructed to reserve their judgement on the comparison weight until they had picked up the standard, but they were unable to refrain from making an immediate judgement using their previous impression of the standard. Their performance was very poor throughout. The direction of the time error was reversed, and the small weights again showed a larger time error than the large weights. For the large weights the PSE was 205·9 gm. and for the small weights 208·1 gm. The difference was insignificant ($p = 0{\cdot}39$ one tail). The standard deviations were larger for both conditions than in Experiment 1, and there was no significant difference between the large and small weight conditions: the SD for the large weights was 17·0 gm. and for the small weights 15·2 gm. Although this difference is in the opposite direction from that of Experiment 1, it can be dismissed as due to chance ($p = 0{\cdot}63$ two tails).

TABLE II

EXPERIMENT 2: COMPARISON FIRST, STANDARD SECOND

| | *PSE* | | *SD* | |
|---|---|---|---|---|
| *Subject* | *Large* | *Small* | *Large* | *Small* |
| 1 | 211·5 | 220·5 | 19·5 | 27·3 |
| 2 | 207·7 | 201·6 | 14·0 | 17·7 |
| 3 | 215·1 | 208·8 | 15·2 | 23·6 |
| 4 | 214·1 | 211·9 | 17·9 | 11·1 |
| 5 | 209·4 | 207·5 | 11·7 | 9·5 |
| 6 | 228·3 | 216·9 | 24·0 | 16·0 |
| 7 | 204·1 | 203·1 | 20·1 | 11·4 |
| 8 | 208·3 | 204·0 | 12·2 | 9·6 |
| 9 | 192·8 | 209·1 | 18·6 | 14·6 |
| 10 | 203·8 | 209·5 | 10·3 | 16·3 |
| 11 | 204·6 | 202·9 | 22·0 | 17·5 |
| 12 | 195·3 | 190·6 | 18·1 | 21·6 |
| 13 | 202·1 | 206·0 | 15·9 | 12·5 |
| 14 | 218·4 | 215·2 | 22·7 | 27·4 |
| 15 | 204·3 | 217·5 | 16·1 | 14·6 |
| 16 | 200·6 | 198·7 | 14·3 | 12·6 |
| Median | 205·9 | 208·1 | 17·0 | 15·2 |

These differences are not significant.

Since this experiment gave results which were too variable, Experiment 3 was undertaken. In this experiment it was hoped to eliminate the constant error, by lifting the weights simultaneously. This was found to be a simpler task than Experiment 2 and provided further useful data.

### EXPERIMENT 3 (Standard and comparison simultaneous)

The procedure was the same as in Experiments 1 and 2, except that the standard and comparison weights were lifted simultaneously, one in each hand. Half the subjects used their dominant and half their non-dominant hand to lift the standard, though hand preference was found to make no difference. It was hoped that if there were no difference in the order of lifting the standard and comparison weights there should be no "time error." Twenty subjects were used.

#### *Results*

This condition gave a result very similar to that of Experiment 1 and gives further evidence that the threshold is higher for the small weights.

The PSE for the large weights was 199·3 gm. and for the small weights 196·0 gm. This difference was not quite significant ($p = 0·068$ one tail), but goes in the same direction as in Experiment 1. That is to say, there is some evidence of a larger constant error for the small weights regardless of the time interval between presentation of the standard and comparison weights.

TABLE III

EXPERIMENT 3: SIMULTANEOUS

| *Subject* | *PSE* | | *SD* | |
|---|---|---|---|---|
| | *Large* | *Small* | *Large* | *Small* |
| 1 | 197·3 | 199·4 | 11·3 | 14·5 |
| 2 | 196·2 | 190·4 | 20·1 | 19·9 |
| 3 | 196·9 | 197·5 | 15·2 | 17·6 |
| 4 | 197·6 | 202·6 | 9·7 | 10·2 |
| 5 | 192·9 | 198·9 | 14·2 | 19·0 |
| 6 | 202·6 | 195·0 | 9·7 | 8·3 |
| 7 | 203·3 | 193·2 | 11·4 | 13·9 |
| 8 | 184·4 | 177·4 | 12·6 | 16·2 |
| 9 | 210·3 | 201·0 | 12·7 | 13·5 |
| 10 | 204·6 | 208·2 | 14·5 | 14·3 |
| 11 | 200·0 | 204·0 | 19·4 | 17·3 |
| 12 | 200·8 | 199·8 | 11·2 | 12·2 |
| 13 | 195·9 | 192·5 | 15·8 | 14·2 |
| 14 | 198·6 | 189·9 | 8·5 | 17·3 |
| 15 | 200·0 | 191·1 | 8·5 | 13·3 |
| 16 | 202·3 | 195·0 | 11·6 | 13·8 |
| 17 | 200·2 | 188·8 | 13·9 | 17·0 |
| 18 | 207·0 | 199·0 | 17·6 | 20·8 |
| 19 | 194·0 | 202·9 | 11·3 | 11·6 |
| 20 | 185·2 | 195·1 | 15·0 | 16·2 |
| Median | 199·3 | 196·0 | 12·6 | 14·4 |

The difference between PSE scores is not significant, $p = 0·068$ (Wilcoxon).
Significance of difference between SD scores, $p < 0·005$ (Wilcoxon).

The SD for the large weights was 12·6 gm. and for the small weights 14·4 gm. The probability of this difference being due to chance was 0·005 (one tail).

## Discussion

We feel justified in dismissing the results of Experiment 2 since the SDs obtained were significantly greater than for either Experiment 1 or 3 ($p < 0{\cdot}02$ on the Kolmogorov-Smirnov test for two independent samples, given in Siegal, 1956, p. 127). Since the SDs and PSEs obtained in Experiment 1 are not significantly different from those obtained in Experiment 3, it seems reasonable to combine the results of these experiments. The combined results give a median SD for the large weights of 12·0 gm. and for the small weights of 14·0 gm. The probability of this difference being due to chance is 0·0003 (one tail). The median PSE for the large weights is 197·0 gm. and for the small weights 194·8 gm. The probability of this difference being due to chance is 0·005 (one tail).

The difference between the constant error for the large and small weights is interesting since Woodrow (1933) found that the direction and size of the "time error" is a function of the physical weight of the standard; heavy standards producing a negative "time error," and light standards a positive one. Our experiment shows that Woodrow's results also apply to the *apparent* weight of the standard; and that constant errors similar to the "time error" are present even when there is no time interval.

It is tempting to think from the direction and size of the threshold change, that $\Delta I/I$ should be regarded as a function of the apparent and not the physical values of I. The size-weight illusion is of the order of 20 per cent. (we cannot give a precise figure, since it fluctuates and tends to lessen during a run) and so, if $\Delta I/I$ were a function of apparent weight, the threshold for the small weights should be raised by about 20 per cent. The observed SD for the large weights was 12·0 gm. and for the small weights 14·0 gm. This is an increase of about 17 per cent. which is fairly close. To establish that Weber's Law should refer to apparent rather than physical intensities would require the measurement of the illusion and of the differential thresholds for many more values.

It would be interesting to discover whether these findings apply only to thresholds for weights, or whether it is a general finding applying to other and possibly all sensory modalities. The question could be answered whenever a consistent scale distortion is present.

This work is supported by U.S. Air Force European Office of Aerospace Research grant No. A.F. EOARDC 61-13.

Two earlier attempts were made to answer the question posed here, by Miss Jean Wallace and Mr. P. Bryant working with the second author, but through no fault of the experimenters their results were not conclusive. Their help is much appreciated. Statistical advice was given by Miss Violet Cane and Mr. E. G. Chambers. We wish to thank Dr. R. Warren for helpful discussions, Dr. R. Audley for most helpful criticism of an earlier draft of this paper, and Professor O. L. Zangwill for advice and encouragement.

## References

Barlow, H. B. (1956). Retinal noise and absolute threshold. *J. opt. Soc. Amer.*, **46**, 634–9.

Fernberger, S. W. (1931). Instructions and the psychophysical limen. *Amer. J. Psychol.*, **43**, 361–76.

Gregory, R. L., and Cane, V. (1955). Noise and the visual threshold. *Nature*, **176**, 1272.

Gregory, R. L. (1956). An experimental treatment of vision as an information source and noisy channel. In Cherry, E. C. (Ed.), *Third London Information Symposium*. London, 201.

Guilford, J. P. (1936). *Psychometric Methods*. New York. 129–31.

Howarth, C. I., and Bulmer, M. G. (1956). Non-random sequences in visual threshold experiments. *Quart. J. exp. Psychol.*, **8**, 163–71.

Seashore, C. E. (1896). Weber's Law in illusions. *Studies from the Yale Psychological Laboratory. IV.*

Siegal, S. (1956). *Non-parametric Statistics for the Behavioural Sciences.* New York.

Tanner, W. P., and Swets, J. A. (1954). A decision-making theory of visual detection. *Psychol. Rev.*, **61**, 401–9.

Woodworth, R. S. (1938). *Experimental Psychology.* New York. 432.

Woodworth, R. S., and Schlosberg, H. (1955). *Experimental Psychology.* London. 205.

Woodrow, H. (1933). Weight discrimination with a varying standard. *Amer. J. Psychol.*, **45**, 391–416.

*Manuscript received 5th October, 1961.*

# TWO TESTS OF ALL-OR-NONE LEARNING AND RETENTION

BY

JOHN BROWN

*From the Department of Psychology, Birkbeck College, University of London*

A number of recent experiments have seemed to imply that the basic associations of learning are formed in an all-or-none manner. However, the evidence provided by these experiments has been essentially indirect since the presence of an association has been assessed by an all-or-none method, viz. simple recall. By allowing the subject more than one attempt at recall or recognition more direct evidence can be obtained. All-or-none learning implies that the probability of success during further attempts will be at the chance level. The experiments described show that this is not the case for the stimulus materials used. On the other hand the uncertainty in bits during further attempts was a fairly high proportion of the uncertainty expected on the all-or-none hypothesis. For this reason the results may indicate only a minor breakdown of the hypothesis. In a retention test administered after a week, fewer first-attempt successes were achieved than in the immediate test but, despite this, the uncertainty after a first-attempt failure was about the same as in the immediate test. The experiments also incorporated a second test of the all-or-none hypothesis. This involved comparing 2-choice and 4-choice recognition scores and recognition scores with recall scores. After corrections for guessing, these scores should be identical, if the hypothesis is correct. Results of this corrected-comparisons test tended to support those of the further-attempts test.

## INTRODUCTION

Considerable interest has been shown recently in the hypothesis that simple associations are learned in all-or-none (AON) manner. Two main tests of this AON hypothesis have been employed by previous investigators. The first can be called the "replacement" test and was devised by Rock (1957). The second can be called the "repeated-recall" test and was first described by Estes (1960).

### *The replacement test*

A list of paired-associates is presented and learning of the pairs is then tested by presenting the stimulus items alone. Any pairs which the subject fails are then replaced by fresh pairs and the procedure is repeated with the modified list. Successive repetitions of the procedure lead to further modified lists until, finally, a complete list is recalled without error. Rock found that with this replacement procedure as few trials were needed to reach a criterion of one error-free trial as with the usual non-replacement procedure. Now replacement of partially learned pairs by fresh pairs, he argued, would make attainment of the criterion more difficult. Accordingly, he concluded that no learning of failed pairs had occurred and hence that learning is all-or-none. Despite its ingenuity, this test has two weaknesses. The first is that equality in trials-to-criterion of the replacement and non-replacement procedures could merely indicate that partial learning does not form a useful basis for complete learning. A second and more serious weakness is that there are a *number* of ways in which the replacement procedure may affect the difficulty of the task. Thus if the favourable effects roughly balance the unfavourable effects Rock's finding would be explained. A number of investigators have followed up this possibility (e.g. Rock and Heimer, 1959; Clark, Dallenbach and Lansford, 1960) but in the writer's opinion interpretation of the replacement test remains problematical. It is suspicious that the replacement procedure sometimes proves to be easier than the non-replacement procedure (Clark *et al.*) and to lead to better retention (Waters and Wogan, 1959).

### *The repeated-recall test*

This test consists in testing twice in succession for the presence of an association by a recall method. Estes (1960) argues that, apart from responses resulting from lucky guessing and from the possibility of some forgetting between the tests, the association will either function in both tests or in neither test, if the AON hypothesis is correct. On the partial learning hypothesis, on the other hand, he considers that successful recall in only one of the tests is quite likely: if an association is at partial strength it will only function sometimes. This test of the AON hypothesis can be criticized on two grounds. First, successful recall in the first test may act as a rehearsal and increase the probability of successful recall in the second test. Second, partial learning need not be interpreted to mean that an association is at partial strength. If it is interpreted to mean that the memory trace carries inadequate information, then partial learning is consistent with a low probability of recall (this interpretation will be further considered in the final discussion). Failure on both tests will then be almost as likely on the partial learning as on the AON hypothesis, i.e. the repeated-recall test is an inadequate method of detecting partial learning.

Both these tests of the AON hypothesis suffer from a common defect. They test for the presence of an association by all-or-none method: on each test occasion the subject either recalls the response or fails to recall the response. Because of this, evidence for or against the AON hypothesis has to depend on what happens subsequently and is necessarily indirect. The merit of the test introduced below is that it does not measure learning in all-or-none manner. In this respect it resembles a test used by Israel (1960).

### *The further-attempts test*

Instead of restricting the subject to a single attempt at recognition or recall, he can be asked to make further attempts when his first attempt is incorrect. If his probability of success on further attempts is better than chance, this indicates that partial learning has occurred. Such a test differs from the repeated-recall test in the following ways: (1) The subject is told whether an attempt is correct. (2) Further attempts are made immediately. (3) The subject can be shown what alternatives he has already chosen so that a different alternative is chosen on each attempt. In the following experiments, subjects were asked to find the correct alternative among a set of alternatives which were equally likely in the absence of learning. Now if attempts after the first have a better-than-chance probability of success, this means that success will tend to occur earlier than would otherwise be expected. Accordingly, the results can be analysed in terms of the probability $p_r$ of success occurring at the rth attempt given that the first attempt has failed. It is easy to show that $p_1 = p_2 \ldots = p_n$ if no learning has occurred. (This method of analysis is more convenient than, but logically equivalent to, an analysis in terms of the probability of success *when making* the rth attempt, since this probability does not remain constant in the absence of learning but rises as successive alternatives are eliminated.)

If the further-attempts test yields statistically significant evidence for partial learning, it is of interest to obtain a measure of the amount of uncertainty as to the correct response when the first attempt has failed. On the AON hypothesis, the expected uncertainty is $\log_2(n-l)$ bits. Upper and lower bounds for his actual uncertainty can be estimated from the observed values of $p_r$ by applying the formulae given below. The ratio of the actual to the expected uncertainty constitutes an inverse measure of the extent to which the AON hypothesis breaks down. The formulae are:

$$\text{Upper bound} = -\sum_{r=2}^{r=n} P_r \log P_r.$$

$$\text{Lower bound} = -\sum_{r=2}^{r=n} (P_r - P_{r+1})\,(r-1)\log(r-1).$$

where $P_r$ is the probability of success occurring at the rth attempt *given that the first attempt has failed* i.e. $P_r = p_r\,(1-p_1)^{-1}$. These formulae are adaptations (cf. Pollack, 1952) of formulae for estimating the redundancy of printed English derived by Shannon (1951).

*The corrected-comparisons method*

This is a second and less direct method of testing the AON hypothesis which was also incorporated in the following experiments. The AON hypothesis implies that the number of first-attempt successes should not depend on the type of test used, once successes attributable to guessing have been eliminated. Thus corrected scores in recognition tests with varying numbers of alternatives and corrected scores in recall tests should all be equal. Corrections for guessing are made on the assumption that the subject either knows the correct response or makes a pure guess. However dubious this assumption may be in general it is entirely appropriate when the hypothesis to be tested is that of all-or-none learning.

## EXPERIMENT I*a*

Estes, Hopkins and Crothers (1960) claimed to get results supporting the hypothesis of AON learning (or, strictly, a slightly modified version of this hypothesis) when a list of paired-associates consisting of 3-consonant stimulus items and single-digit response items were presented at a rate of 8 sec. per pair. It was therefore decided to see whether evidence obtained by using the further-attempts and corrected-comparisons tests would also accord with the AON hypothesis. A single presentation of the list was followed by either a 2-choice or a 4-choice test condition during which the stimulus items alone were presented. The subject was required to find the correct response to each stimulus item among two alternatives in the 2-choice condition and among four alternatives in the 4-choice condition.

### METHOD

*Lists*

An N–L and an L–N list each containing ten pairs, were prepared. The N–L list was intended as a practice list and consisted of 3-digit stimulus items (digits from 1 to 9) followed by one of the four letters A, B, C, D, as the response item. The L–N list consisted of 3-consonant stimulus items followed by one of the four digits 1, 2, 3, 4, as the response item. The stimulus items were selected randomly without constraints and the response items were selected randomly but with the constraint that no item was used more than three times. The object of this constraint was to minimize the effect of any response bias in favour of one of the letters or digits.

*Response boards*

Each board consisted of a layer of hardboard fixed to ½ in. chipboard. The alternatives were duplicated on a sheet of paper (the test-sheet) which covered the board. Next to each alternative on the sheet was a small circle. The circle was directly above the hole in the board. The hole was deep for a correct and shallow for an incorrect alternative. The subject pressed through the paper with a stylus in order to determine the depth of the hole and in doing so left a record of his attempt. The stylus consisted of a small electrical screwdriver cut to such a length that the whole of the shaft would disappear down a correct but not an incorrect hole. For the 4-choice condition the test sheet contained ten rows of A B C D in the left half and ten rows of 1 2 3 4 in the right half. The 2-choice test sheet was formed by omitting two alternatives from each row in such a way that the correct alternative was on the right and the left equally often.

### *The presentation apparatus*

Owing to the fact that subjects needed to see their response boards, it was necessary for presentation to take place with the room well illuminated. Projection methods were therefore discarded in favour of direct viewing of cards on which letters and numbers 2 in. high were stencilled, the stimulus item in black and the response item in red. An unorthodox but effective card-changer was constructed in which the cards hung one behind another on separate fuse wires. These fuse wires were then blown from the front at 8-sec. intervals by an appropriately connected uniselector and timer. As each fuse blew, the card hanging from it dropped smartly from sight leaving the next card exposed.

### *Procedure*

At the outset of the experiment the nature of the task was explained and the composition of the stimulus and response items of the first list was stated. The ten pairs of the N–L list were then presented one at a time at a rate of 8 sec. per pair. After an interval of 8 sec., the stimulus items were presented alone at the same rate but in a different order: subjects had been warned of this change in the initial description of the task. As each stimulus item was presented, the subject attempted to find the correct alternative on his response board and continued to make attempts until he succeeded. The first row of alternatives related to the first stimulus item, the second row to the second stimulus item and so on. The procedure was repeated with the L–N list after an interval of several minutes during which the composition of the items of this list was stated.

### *Subjects*

The subjects were university students and were tested in three separate groups of 18, 15 and 28. The subjects within each group were allocated to the two conditions at random. In addition, 5 subjects participated in the experiment but failed to carry out the instructions correctly with either list.

## RESULTS

The 4-choice results will be considered first. Table I shows the probabilities of success occurring at the 2nd, 3rd and 4th attempts ($p_2$, $p_3$ and $p_4$ respectively) obtained by averaging over all subjects. On the AON hypothesis we expect $p_2 = p_3 = p_4$. In Table I, $p_2$ is greater than $p_3$ or $p_4$ for both lists and $p_3$ is virtually equal to $p_4$ for the N–L list and greater than $p_4$ for the L–N list. These trends are in the direction expected on the partial-learning hypothesis. Since $p_2$ and $p_4$ are expected to differ most on this hypothesis, a simple test of the significance of each trend can be obtained by examining the relation of $p_2$ to $p_4$ for individual subjects.

TABLE I

PROBABILITIES OF SUCCESS OCCURRING ON THE FIRST, SECOND, THIRD AND FOURTH ATTEMPTS

| | *Expt. Ia* | | *Expt. Ib* | | *Expt. II* | | *Expt. II* | |
|---|---|---|---|---|---|---|---|---|
| | *N–L* | *L–N* | *N–L* | *L–N* | *Words I* | *Nos. I* | *Words II* | *Nos. II* |
| $p_1$ | 0·413 | 0·532 | 0·559 | 0·579 | 0·868 | 0·708 | 0·705 | 0·588 |
| $p_2$ | 0·226 | 0·200 | 0·245 | 0·259 | 0·085 | 0·185 | 0·195 | 0·233 |
| $p_3$ | 0·178 | 0·168 | 0·197 | 0·162 | 0·032 | 0·056 | 0·065 | 0·116 |
| $p_4$ | 0·183 | 0·100 | — | — | 0·016 | 0·051 | 0·036 | 0·062 |

Consider the group of subjects for whom $p_2 > p_4$ and the group for whom $p_2 < p_4$. On the AON hypothesis we expect equal numbers of subjects in these two groups. On the partial learning hypothesis we expect more subjects in the group for whom $p_2 > p_4$. (Moreover, since this is the only alternative hypothesis, a one-tail test of significance might well be justified.) The number of subjects in these two groups did not differ significantly for the N–L list ($0\cdot05 < p < 0\cdot1$) but did differ significantly for the L–N list ($p < 0\cdot01$) on two-tail tests. Table II shows the upper and lower bounds for the average uncertainty shown by subjects after failing on their first attempt: in both cases the expected uncertainty on the AON hypothesis is $\log_2 3$, i.e. 1·585 bits. Assessed in this way the results do not depart much from what is expected on the AON hypothesis, since even the lower bounds are 80–90 per cent. of the expected uncertainty.

TABLE II

UPPER AND LOWER BOUNDS FOR THE UNCERTAINTY AFTER FAILING THE FIRST RESPONSE

| | *Expt. Ia* | | *Expt. Ib* | | *Expt. II* | | *Expt. II* | |
|---|---|---|---|---|---|---|---|---|
| | *N–L* | *L–N* | *N–L* | *L–N* | *Words I* | *Nos. I* | *Words II* | *Nos. II* |
| UB | 1·576 | 1·531 | 0·991 | 0·961 | 1·275 | 1·314 | 1·246 | 1·381 |
| LB | 1·456 | 1·310 | 0·890 | 0·790 | 0·810 | 0·866 | 0·776 | 0·977 |
| E | 1·585 | 1·585 | 1·000 | 1·000 | 1·585 | 1·585 | 1·585 | 1·585 |

UB = upper bound, LB = lower bound, E = uncertainty expected on AON hypothesis. Uncertainties expressed in bits.

The comparison of corrected scores was confined to the L–N list since, in the case of the practice list N–L, a number of subjects failed to continue their attempts until the correct alternative was found on every row. Scores were corrected for guessing by subtracting the number of incorrect first-attempts divided by (n–l) from the number of correct first-attempts: this is appropriate because in a test with n alternatives an unlucky guess is (n–l) times as likely as a lucky guess if learning has been all-or-none. The corrected 2-choice and 4-choice scores were compared by the Mann-Whitney non-parametric test (the distribution of scores being non-normal) and were found not to differ significantly. However, there was considerable variability between subjects and the 2-choice mean was 4·8 whereas the 4-choice mean was only 3·8.

The 4-choice results appear to provide significant evidence for at least some partial learning. However, 4-choice subjects had up to four alternatives to examine and test in the 8-sec. interval between successive stimulus items. Since a number of subjects failed to find the correct alternative on a number of rows of the response sheet, this suggests that 4-choice subjects were performing under an undue time stress. If haste can lead to the selection of the wrong alternative on the first attempt, this might explain the greater frequency of second attempt successes as compared with third and fourth. It was therefore decided to perform a further experiment in which only three alternatives were provided on each row of the response sheet and in which subjects were instructed to make only *two* attempts to find the correct alternative: here it is still possible to determine whether second-attempt successes are more numerous than they should be on the AON hypothesis.

The time stress under which 4-choice subjects were apparently working affects the validity of comparing corrected 2-choice and 4-choice scores. In any case, the comparison was inconclusive since no significant difference was found despite a moderately large difference between means.

## EXPERIMENT Ib

In this experiment a further group of 29 university students were tested under a 3-choice condition. They were instructed to make a second attempt to find the correct alternative on each row of the response board, if their first attempt failed, but not to make a third attempt. The same lists were used as in the previous experiment. The paper sheets covering the response boards carried only three alternatives per row: the first, second and third alternatives were correct on at least 3 out of the 10 rows.

All subjects performed the task according to the instructions on all rows with both lists. This suggests that they were not working under an undue time stress. Table I shows that $p_2$ was higher than $p_3$ for both lists. Significantly more subjects succeeded more often on their second attempt than on their third attempt ($p < 0.05$ on $\chi^2$ tests), as expected on the partial learning hypothesis. After a first-attempt failure the expected uncertainty on the AON hypothesis is $\log_2 2$ or 1 bit. Table II shows that both the upper and lower bounds for the actual uncertainty are not much below this expected uncertainty.

## EXPERIMENT II

Estes (1960) briefly described an experiment on repeated recall of words and claimed that the results supported the AON hypothesis. The free-recall method was used, i.e. the subjects were not required to recall the words in the original order. Thus the type of association presumed to be learned in the AON manner is the association that a word belongs to a recently presented group of words. In the following experiment, the learning of associations of this type was tested by 2-choice and 4-choice recognition tests, analogous to those used in Experiment I, and also by a free-recall test. Again, examination of the 4-choice results provides a test of AON learning by the further-attempts method. Comparison of all three sets of scores, after correction for guessing, provides a test of AON learning by the corrected-comparisons method. Both the recognition tests and the recall tests were repeated a week later with a switch-round of subjects to provide a counter-balanced design. The results in these delayed tests should show increased evidence for partial learning if learning becomes partial through trace-decay, as will be suggested in the final discussion.

Two lists were used, a list of words and a list of numbers. It should be mentioned that the experiment was partly designed to make possible an information analysis which it is hoped to present elsewhere.

## METHOD

### *Design*

For the immediate test, subjects were allocated randomly to the three conditions, 2-choice and 4-choice recognition and recall. Those subjects who performed under any one condition for the immediate test were again allocated randomly to the three conditions for the delayed test. This was intended to guard against the possibility that retention was affected by the nature of the immediate test. Three separate groups of subjects were tested and it was arranged that approximately equal numbers of subjects served under each of the conditions in both the immediate and delayed tests: this constituted the only constraint on random allocation.

### *Procedure*

At the outset of the experiment the nature of the three test conditions was explained. In particular, subjects were told (1) that there was always just one correct alternative on each row and that they should continue to make attempts until this alternative was found, (2) that in the recall condition the order in which the words were written did not matter, (3) that the list would consist of 16 consonant-vowel-consonant 3-letter words. The word-list was then presented at a rate of 8 sec. per word. As soon as presentation was complete, recognition subjects removed a flap which had until then covered the rows of alternatives on the response boards in front of them. The immediate test then commenced and lasted for 2½ min. When the test was finished recognition subjects replaced the flaps so that the alternatives were again covered and recall subjects turned over the sheets of paper on which they had been writing. The same procedure was then repeated with the number-list except that the description of the test conditions was omitted. The possibility of a delayed test was not mentioned. The delayed test, which also lasted 2½ min. for each list, was given one week later (without any further presentation of the lists).

### *Lists and test sheets*

*Words.* All consonant-vowel-consonant words in the English language were enumerated by trying the five vowels in all possible pairs of consonants. The number of combinations recognized as words was 284. After dubious and less-known words had been eliminated, 256 remained from which a list of 16 words was randomly selected. This somewhat elaborate method of obtaining the list was designed to facilitate information calculations. Two 4-choice test sheets (one for the immediate test and one for the delayed test) were constructed as follows: each sheet contained 16 rows and each row consisted of one word from the list and of three new words from the ensemble of 256 words. The order of the words in each row was random except that an equal number of correct words was made to occur in each column. Differences between the two sheets were as follows: (1) The pattern of correct holes differed (2) the rth row in each sheet contained a different correct word, (3) half the words from the list were presented alongside the same incorrect words in both sheets and half with different, incorrect words, rows of these two types alternating. The two 2-choice sheets were identical with the 4-choice sheets except that two alternatives were omitted in each row in such a way that the alternative on the right and the alternative on the left was correct equally often. *Numbers.* The list of 16 numbers was randomly selected from 90 2-digit numbers between 10 and 99. The test sheets were constructed according to the same principles as the test sheets for the word list.

### *Subjects*

The subjects were 85 university students of whom two failed to carry out the instructions correctly and were dropped from the experiment. Of the remaining 83 subjects, 11 were absent for the delayed test which, as stated previously, was administered without prior warning.

## RESULTS AND DISCUSSION

The results from the immediate test (Words I and Numbers I) will be presented first. Table I shows that with the 4-choice subjects there was a trend in the direction $p_2 > p_3 > p_4$, for both lists, as expected on the partial learning hypothesis. In each case the trend was significant at the 0·01 level when assessed as in Experiments I*a* and I*b*. Table II shows the upper and lower bounds for the uncertainty displayed by subjects after failing on their first attempt. It will be seen that less uncertainty was displayed than in Experiments I*a* and I*b*, i.e. there was more evidence of partial learning. Table III shows the mean percentage scores corrected for guessing under the 2-choice, 4-choice and recall conditions. It also shows the significance of the difference between each pair of conditions as assessed by the Mann-Whitney test (as before this test was selected because of non-normality in the scores). Contrary to what is expected on the AON hypothesis there are significant differences between the means. The main difference is between the 2-choice means on the one hand and the 4-choice and recall means on the other. A possible explanation for this finding is that the 2-choice test takes less time to perform than the 4-choice test or the recall test, since forgetting during the time it takes to perform each test is not unlikely. As would be expected if this explanation is correct, it was found that scores for the top halves of the recognition sheets tended to be higher (about six percentage points) than for the bottom halves. An unexpected finding was that the order of the 4-choice and recall means differed for the two lists (Table III). However, spontaneous remarks by subjects suggested that the numbers were often coded into systems during presentation and presumably it is easier to make use of such coded information in the recall test since in the recognition test the order of recognition is largely imposed.

TABLE III

MEAN PERCENTAGE SCORES CORRECTED FOR GUESSING

| | *a*<br>*2-choice* | *b*<br>*4-choice* | *c*<br>*recall* | *Significance of comparisons* | | |
|---|---|---|---|---|---|---|
| | | | | *a–b* | *a–c* | *b–c* |
| Words I .. | 95·6 | 82·3 | 67·2 | < 0·01 | < 0·01 | < 0·01 |
| Words II .. | 70·3 | 60·7 | 19·7 | N.S. | < 0·01 | < 0·01 |
| Nos. I .. | 80·6 | 61·1 | 67·6 | < 0·01 | < 0·01 | N.S. |
| Nos. II .. | 50·5 | 45·2 | 33·7 | N.S. | < 0·05 | N.S. |

The results of the delayed test (Words II, Numbers II) are similar to those for the immediate test. Again there was a trend in the direction $p_2 > p_3 > p_4$ (Table I) which was found to be significant at the 0·01 level for both lists and again there were significant differences between corrected scores (Table III). This is not surprising since partial retention is an almost inevitable sequel to partial learning. It will be seen from Table II that the average uncertainty after failing the first response (the conditional uncertainty) was about the same in the immediate and delayed tests. Now considerably fewer first-attempt successes were achieved in the delayed test and this implies that the *total* uncertainty was considerably greater in this test. Thus in relation to the total uncertainty the conditional uncertainty is *lower* in the delayed test. This accords with the hypothesis that partial learning is a product of trace-decay since after a week there will be fewer traces from which no information loss has occurred.

### *Subsidiary results*

In the delayed recognition tests, one half of the items were presented among the same wrong alternatives as in the immediate test and the other half among new wrong alternatives. Any partial learning of wrong alternatives during the performance of the immediate test might be expected to lead to confusion when the same wrong alternatives were used in the delayed test. However, no evidence for such an effect could be discerned in the results: in so far as a trend was present, it was in the direction of *more* accurate recognition when the same wrong alternatives were used. The results were also examined to see whether the nature of the immediate test affected performance in the delayed test. It was found that recall of both words and numbers was about twice as high for those subjects who had also performed under the recall condition in the immediate test. Mann-Whitney tests comparing these subjects with the remaining subjects in the delayed test confirmed that the improvement in performance was significant for both the number list and the word-list at the 0·05 level. This finding is reminiscent of the claim that recitation is an aid during memorizing (Gates, 1917). In no other instances did the nature of the immediate test appear to affect performance in the delayed test.

## DISCUSSION

On the hypothesis of AON learning, successes after the first attempt should occur on a chance basis. In all the experiments there was clear evidence that this was not the case. A similar finding was obtained by Pollack (1952). In an experiment on information transmission, he found that attempts to correct erroneous digits in the immediate reproduction of a sequence of digits had a better than chance probability of success. However, his results are somewhat equivocal since no account was taken of whether errors on the first attempt represented transposition errors and correction of transposition errors, even on a chance basis, would lead to better than chance performance in correcting individual digits. In the foregoing experiments, subjects were not required to reproduce the order of the items so that the problems raised by the occurrence of transposition errors were avoided.

*Prima facie*, the *extent* to which the further-attempts results departed from those expected on the AON hypothesis was relatively small in terms of the information measure. Thus in the paired-associates experiments, the uncertainty as to the correct response after a first-attempt failure was 79–99 per cent. of the expected uncertainty and, in the remaining experiment, the uncertainty was 51–83 per cent. of the expected uncertainty. (These percentages are derived from Table II by taking the lowest lower-bound estimate and the highest upper-bound estimate). On the other hand, it is important to recognize that when responses are partially learned they contribute to first-attempt successes as well as to successes after the first attempt. Thus the fact that the results did not depart dramatically from those expected on the AON hypothesis does not necessarily mean that the hypothesis is substantially correct. A more definite conclusion would be possible if information could be obtained on how the observed distribution of $p_1$, $p_2$, $p_3$ and $p_4$ is generated. It might be possible to obtain such information by asking the subject to indicate which of the alternatives he regards as possible before he commences his attempts to find the correct alternative with the stylus. If, for example, he is always correct when he indicates only one alternative as possible, this would show that complete learning had occurred in relation to the items concerned.

In Experiment II, the results of the corrected-comparisons test also provided evidence against the hypothesis of AON learning. Although the difference between the corrected 2-choice and 4-choice scores in Experiment I*a* was not significant statistically, it was in the same direction as in Experiment II. According to a probability model which it is hoped to present elsewhere,* corrected 2-choice scores are necessarily higher than corrected 4-choice scores when $p_2 > p_3 \geqslant p_4$, as was the case in both experiments. This implies that any explanation of the results of the further-attempts test is potentially an explanation of the results of the corrected comparisons test as well.

Both tests showed some deviation from the results expected on the AON hypothesis. The question arises of whether these deviations can be explained in such a way as to leave the hypothesis intact. Various possible explanations will now be considered.

---

* The model is described in a paper entitled "Multiple response evaluation of discrimination" which will appear in *Brit. J. statist. Psychol.*

### *Unsuitability of the units*

The units employed in the experiments were chosen because previous investigators using the same or similar units have claimed to get results supporting the AON hypothesis. Even if it is true that individual associations are learned in an AON manner, partial learning of a complex unit will be possible since it will involve a number of associations each of which may be either learned or not learned. It is difficult to think of any unit which is not potentially complex in this way. Even when a unit is perceived as a unit through prior familiarity, some of its attributes may be noted separately. Now if possession of a particular attribute is learned but not the unit as such, this will usually constitute *partial* learning of the unit. For example, the digit "2" is a familiar unit. Nevertheless, if the subject can only remember that the correct digit is even, this will represent partial learning.

### *Non-random response tendencies*

Even if responses after the first are unguided by partial learning, it is unlikely that they will be made randomly in the strict sense. Non-random response tendencies would affect the distribution of successes if they happened to link up with the pattern of correct holes on the response board. Evidence that this did not happen to any important extent is provided by the fact that similar results were obtained when different patterns of correct holes were used. However the constraint that each column on the response board contained an approximately equal number of correct holes could have biased the results with all patterns, since subjects may have imposed this constraint on their first choices. As a check on this, some of the results for the first and second halves of the boards were analysed separately but the trend towards second attempt as opposed to third or fourth attempt successes was as marked for the first half as the second half.

### *Generalization*

If one of the presented alternatives in the recognition test is similar to the correct alternative, it might be selected through a generalization process. However, it is necessary to explain why it is preferred to the correct alternative. Indeed, it seems reasonable to regard such generalization as a consequence of partial as opposed to complete learning.

The deviations from what is expected on the AON hypothesis are not necessarily to be explained away. Alternative interpretations of the results should be considered. Results apparently incompatible with a simple AON hypothesis have also been obtained by other workers (Estes, Hopkins and Crothers, 1960; Israel, 1960; Jones, 1962). One possibility is to modify the AON hypothesis by supposing that the stimulus on a test trial differs somewhat from that used on a training trial so that a response, although learned, may not be elicited (Estes *et al.*, 1960). Another possibility is to interpret the results in terms of the Decay Theory of memory.

On the Decay Theory of memory, as elaborated by Brown (1959), every perception leads to the establishment of a memory trace representing the information contained in the perception. Thus at the outset learning is always complete. However, the trace undergoes random deterioration with the passage of time (decay). The rate of decay is the same for all traces and is initially high but then falls. Whether decay leads to information loss (forgetting) depends not only on the amount of decay but also on the amount of redundancy present when the trace is formed. Roughly speaking, redundancy implies multiple representation of the information so that information lost from one part of the trace may still be present elsewhere in the trace. On this theory, partial learning is always possible in the sense that, at the time learning is tested, some of the information originally stored in the trace may have been lost through decay while some still remains. This applies even when the information stored is very simple. For example, suppose that the trace is merely required to represent that one of two equally likely alternatives is correct (one "bit" of information). Decay could leave the trace in a state which failed to define the correct alternative but which made one alternative more probable than the other. This theory clearly accords with the finding that further-attempts have a better than chance probability of success and also with the finding that partial learning forms a greater proportion of the total learning a week after presentation than immediately after. On the other hand, the finding that the uncertainty after a first-attempt failure was a high proportion of the maximum possible uncertainty is apt to seem surprising if the theory is correct. However, as already pointed out, partial learning contributes to first-attempt successes as well as to successes after the first attempt so that the implication of this finding is at present unclear.

## REFERENCES

BROWN, J. (1959). Information, redundancy and decay of the memory trace. In *The Mechanization of Thought Processes*. Nat. Phys. Lab. Sympos., No. 10. H.M.S.O.

CLARK, L. L., DALLENBACH, K. M., and LANSFORD, T. G. (1960). Repetition in associative learning. *Amer. J. Psychol.*, **73**, 22–40.

ESTES, W. K. (1960). Learning theory and the new "mental chemistry." *Psychol. Rev.*, **67**, 207–23.

ESTES, W. K., HOPKINS, B. L., and CROTHERS, E. J. (1960). All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, **60**, 329–39.

GATES, A. I. (1917). Recitation as a factor in memorising. *Arch. Psychol., N.Y.*, **6**, No. 40.

ISRAEL, M. L. (1960). Variably blurred prompting: I. Methodology and application to the analysis of paired associate learning. *J. Psychol.*, **50**, 43–52.

JONES, J. E. (1962). All-or-none versus incremental learning. *Psychol. Rev.*, **69**, 156–60.

POLLACK, I. (1952). The informational contribution of "wrong" responses. *H.R.R.L. Memo Report No. 25.*

ROCK, I. (1957). The role of repetition in associative learning. *Amer. J. Psychol.*, **70**, 186–93.

ROCK, I., and HEIMER, W. (1959). Further evidence of one-trial associative learning. *Amer J. Psychol.*, **72**, 1–16.

SHANNON, C. E. (1951). Prediction and entropy of printed English. *Bell. Syst. tech. J.*, **30**, 50–64.

WATERS, R. H., and WOGAN, M. (1959). The role of repetition in associative learning. *Amer. J. Psychol.*, **72**, 612–3.

*Manuscript received 8th October,* 1962.

# REDUNDANCY IN NON-SEQUENTIAL PATTERNS

BY

ALAN STANILAND

*From the Psychological Laboratory, University of Exeter*

Recognition, discrimination and delayed discrimination tasks were carried out with selections of composite visual patterns, ordered in terms of a generalized measure of redundancy based upon transmission between determinates under four determinables. Results in the recognition experiment showed a significant relationship with redundancy and there was indication of a similar effect in delayed discrimination. Redundancy levels, however, appeared irrelevant to simultaneous discrimination, where each response could be made discretely without reference to the structure of the selection as a whole.

## INTRODUCTION

An earlier paper (Staniland, 1960) discussed certain aspects of redundancy which are shown in a complete table of decomposition of probabilities. It was suggested that such a table displays variations of structure which are likely to be relevant to performance, but to which Shannon's frequently used measure of redundancy for digit sequences is not sensitive. The essential problem is to state a measure by which selections of equal numbers of items can be legitimately regarded as nevertheless of different redundancy, but in the earlier paper the methods used by Fitts and Miller, depending upon random or non-random methods of selection, were criticized. The writer's chief concern, out of which the earlier paper arose incidentally, had been the possibility of applying a redundancy measure to selections of visual patterns for which no fixed sequence of scanning could be assumed, and of which the constituents were not confined to a strictly limited alphabet. It was from further consideration of the results of decomposition that a variant measure of redundancy was adopted which, within certain limits, could be applied to non-sequential patterns. The mathematical argument is given briefly, and the remainder of the paper reports upon its experimental use in tasks of recognition and simultaneous and delayed discrimination.

Redundancy in visual forms has two fairly distinct senses. One, especially considered by Attneave (1954), concerns the internal conditionalities which may be found within a single pattern—symmetry, continuity of outline, etc. The other refers to the redundancy of sets of visual forms which can be regarded as drawn, either randomly or by some principle of selection, from a definable theoretical ensemble. This aspect has been particularly investigated by Fitts and his associates (1956 and later). Work by Miller (e.g. 1958) can also be cited, because though strings of letters were used they were based upon a finite-state generator and raise issues that might also be relevant to spatial patterning in general. The present paper follows the second approach, but has a certain relationship with the first in that it is concerned not only with the method of selection and gross Information content of a set of patterns, but also with those features of the internal structure of the set which can only be determined by a more detailed analysis of each specific selection. In the terms used by Garner in a recent book (1962), it is concerned not only with the amount but also the form of redundancy.

In the earlier paper, redundancy was discussed in relation to selections of $n$ items, each item consisting of a sequence of $c$ digits drawn from $s$ types of digit or symbol. In such a selection the average Information gain over all first digits was referred to as the gain on the first "step," the average gain over all second digits as the gain on the second step, and so on. The upper limit to Information at any step is the uncertainty of the digits at the step $(\Sigma - p \log_2 p)$ and this will be actual average gain at the first step. But at subsequent steps redundancy will show itself as a reduction of the actual gain compared with the uncertainty of digits, and the gain tends to reduce step by step until the maximum actual gain ($\log_2 n$) is obtained. It was suggested that the manner of this reduction was indicative of features of a selection which the standard measure $[1 - (\log_2 n / c \log_2 s)]$ fails to distinguish, but which may be relevant to psychological performance.

The question here considered is the extension of a similar approach to selections of visual patterns which can be regarded as partially analogous to digit codings insofar as they are composites of distinct characteristics: colour, form, size, etc. These differ from digit codings, in two important respects:

(i) A regular order of scanning cannot be plausibly assumed.

(ii) Similar characteristics cannot be said to recur as alternatives in the way in which, e.g., the letter "A" may occur either in the first or any other position. The assumption of such recurrence is built into the standard formula as used by Fitts and Miller.

A more generalized measure can be approached by considering that a simple coding such as AAB is in effect equivalent to $A_1A_2B_3$, in which the subscripts take over the positional indication, so that, e.g., $A_2B_3A_1$ is identifiable with $B_3A_1A_2$ or any other disposition of the "position-free" (but now complex) symbols. (This point is raised also by Pierce (Cherry, 1961) in a discussion-note on a paper by Crossman). Once this aspect of regular sequence is recognized, it is apparent that the *same* symbol does not, in one sense, re-occur in a different position, but that what is strictly relevant is the variety and probabilities of symbols in each group of alternatives however the group is identified. In the sequential-matrix each column forms such a group (the group of "first digits," the group of "second digits," etc.) and its composition is summed up in the column uncertainty ($H_1$). In other selections, characteristics other than sequence may mark the groups of alternatives and a useful terminology is that first suggested by Johnson (1921) for discussing variations of qualities within a related range of qualities. Instead of "symbols in the same column," we can generalize to "determinates under the same determinable," determinables being what might be called categoric dimensions of quality. The use of the terminology is exemplified by variations which could occur in a selection of simple patterns:

| *Determinables:* | *Shape* | *Colour* | *Height* |
|---|---|---|---|
| Determinates under each determinable | Square | Red | 1 in. |
| | Circular | Blue | 2 in. |
| | Triangular | Green | 3 in. |

## *A generalized measure*

Returning to the conventional type of selection, we can now consider the possibility of generalization. For the selection tabulated in the 1960 paper ($c = 4$; $s = 4$; $n = 24$) the step gains ($h_1$) were as follows:

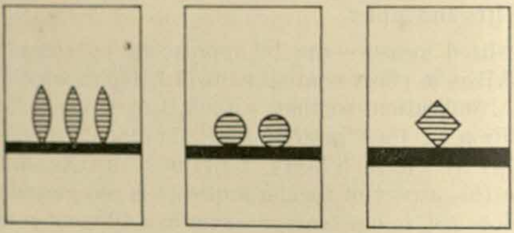
| Step: | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| Actual Inf. gain ($h_1$) .. | 1·667 | 1·458 | 0·650 | 0·810 | $\Sigma h_1 = 4·585 = \log_2 24$ |
| Information available ($H_1$) | 1·667 | 1·959 | 1·855 | 1·979 | $\Sigma H_1 = 7·460$ |
| $T_{1 \ldots c}$ .. | 0 | 0·501 | 1·205 | 1·169 | $\Sigma T_{1 \ldots c} = 2·875$ |
| | | | | | $[7·460 - 2·875 = 4·585]$ |

The Information gain at the first step ($h_1$) is, of course, equal to the Information available ($H_1$), i.e. $\Sigma - p \log_2 p$ for the symbol probabilities at the first position over all items. At subsequent steps, where the probabilities are limited by the already ascertained digits, $h_1$ is less than $H_1$, and this reduction can be shown to be equal at each step to the transmission between already ascertained digits and the digits at that step. Thus $h_2 = H_2 - T_{1:2}$; $h_3 = H_3 - T_{1,2:3}$; and $h_c = H_c - T_{1,(c-1):c}$. If we represent the sum of these losses as $\Sigma T_{1\,..\,c}$ (or for brevity $\Sigma T$ (see mathematical note given as appendix)), we have, therefore:

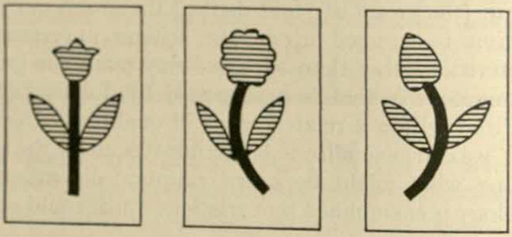
$$\log_2 n = \Sigma h_1 = \Sigma H_1 - \Sigma T_{1\,..\,c}$$

Although the composition of $\Sigma T$ may vary if the order of columns is changed, as a sum it must be independent of order, i.e. if columns are permuted as columns, retaining the vertical order of symbols, each permutation gives a new selection which equally discriminates *n* items and gives $\log_2 n$ average Information per item. $\Sigma T$ thus represents Information necessarily redundant from the constitution of the columns, irrespective of order. Comparing $\Sigma T$ (the Information wasted) with the Information potentially available gives the ratio:

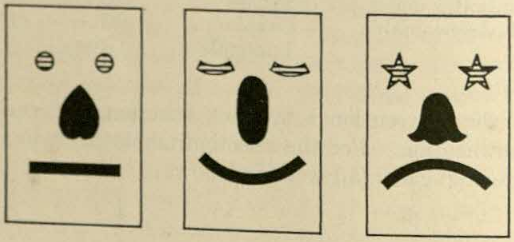
FIGURE 1



Type A



Type B



Type C

Specimens of display material. All the shapes used appear in these sample patterns. Shaded parts varied in colour: black parts only in form.

$$\frac{\Sigma T}{\Sigma H_1} = \frac{\Sigma H_1 - \log_2 n}{\Sigma H_1} = 1 - \frac{\log_2 n}{\Sigma H_1}$$

The similarity of this expression to Shannon's measure of redundancy (Shannon, 1940) is obvious, and the two are in fact equal when Shannon's redundancy is zero, for then ΣT is also zero. But this measure assesses the utilization of symbols apart from the assumption that the same symbols are available at all steps. It states directly that a selection is redundant or non-redundant according to the degree of transmission between columns, and it includes Shannon's measure as a special (and for *his* problem, justifiable) case.

We shall refer to the suggested measure as D (decompositional) redundancy, since it can be calculated only by analysis which takes account of the detailed disposition of symbols, or, in its more general application, of the range of determinates under each determinable. Shannon's measure gives (on the model of an exponential curve with *c* as length, *s* as exponent) the wasted length of symbolization: D-redundancy makes a not dissimilar criticism of efficiency of coding, comparing the number of items actually formed with the number that could have been formed with the most efficient disposition of symbols and a minimum ΣT. It can take more values than Shannon's measure and is, therefore, potentially more discriminating. It is by this measure that redundancies are assessed in all the experiments that follow, and its application will be demonstrated in relation to the sets of patterns which were used.

### *Display material*

The experiments which follow are based upon selections of nine patterns of one of three types, and these are conveniently described first. The basic formal structure of each set of patterns was three-way variations of four characteristics, giving a theoretical ensemble of $3^4$ or 81 patterns. This was the structure of the pack used by Bruner (Bruner, Goodnow and Austin, 1956) for experiments in categorization, and a pack of this type was used for brief preliminary experiments and also to obtain random selections of nine patterns of varying redundancies. The selections used had percentage redundancies of: 0; 6; 19·8; 24·5; 43·6; 45·3; forming three pairs of Low, Medium and High redundancy. Sample selections are given in Table I. The three types (Fig. 1) were:

A. An abstract pattern, modified from the Bruner pack.
B. A representational natural unit, a flower with two leaves.
C. Made up of disconnected shapes but with a very obvious face gestalt.

They were given the redundancy structures in terms of characteristics 1,2,3 in columns (determinables) w,x,y,z, as follows:

| Determinables: | w | x | y | z |
|---|---|---|---|---|
| Type A | Depth of band (1, 2 or 3 units) | No. of symbols (1, 2 or 3) | Symbol colour (red, blue or brown) | Symbol shape (circle, petal or diamond) |
| Type B | Stem shape (straight, or curved right or left) | Leaf colour (green, yellow or brown) | Flower colour (red, yellow or blue) | Flower shape (round, heart or bell) |
| Type C | Mouth shape (straight or curved up or down.) | Nose shape (bell, inverted heart or oval) | Eye shape (round, curved or star) | Eye colour (blue, brown or green) |

The patterns were made up of coloured paper shapes (sold for children's play) which were machine-cut and constant in shape and colour. These were used (for Experiment 1) stuck on to white cards of normal playing-card size, or (for Experiments 2 and 3) photographed to give colour transparencies.

TABLE I

SAMPLE SELECTIONS

*Medium Redundancy* (19·8 *per cent.*)

|  | *w* | *x* | *y* | *z* |
|---|---|---|---|---|
|  | 2 | 1 | 2 | 2 |
|  | 2 | 2 | 2 | 2 |
|  | 2 | 3 | 2 | 2 |
|  | 2 | 1 | 2 | 3 |
|  | 2 | 2 | 2 | 3 |
|  | 2 | 3 | 2 | 3 |
|  | 2 | 2 | 3 | 3 |
|  | 3 | 2 | 2 | 2 |
|  | 3 | 2 | 3 | 2 |
| $H_1$ | 0·764 | 1·434 | 0·764 | 0·991 |

$\Sigma H_1 = 3{\cdot}953$

*High Redundancy* (43·6 *per cent.*)

|  | *w* | *x* | *y* | *z* |
|---|---|---|---|---|
|  | 2 | 2 | 1 | 2 |
|  | 1 | 2 | 2 | 1 |
|  | 1 | 2 | 3 | 1 |
|  | 2 | 3 | 3 | 1 |
|  | 2 | 1 | 3 | 2 |
|  | 2 | 3 | 3 | 2 |
|  | 2 | 1 | 3 | 3 |
|  | 2 | 3 | 3 | 3 |
|  | 3 | 3 | 2 | 2 |
| $H_1$ | 1·351 | 1·530 | 1·224 | 1·530 |

$\Sigma H_1 = 5{\cdot}635$

## EXPERIMENT 1. RECOGNITION

For a first experimental test of D-redundancy the chief decision to be made was the kind of task to be investigated and this was largely determined by the fact that since the material did not use fully-established units, such as letters or digits, it was unsuitable for reproductive recall. Recognition was preferred to discrimination because it demanded storage—and presumably structuring—of the whole set of displays. There was however a difficulty in the standard form of recognition task in which already-displayed items are repeated among novel items: it would involve the redundancy not only of the original set, but also the redundancy of the whole set in the recognition trial. To avoid this complication a form of progressive recognition was designed in which only a single set of displays appeared.

*Procedure.* Twenty four subjects were tested (11 males, 13 females, mostly students of psychology, music, drama and art.). Each had three trials, in each of which he was presented with 36 cards, bearing nine patterns of the same type, each of which was presented four times. The subject was asked to indicate for each card whether or not it had been presented before. For any one subject, one of the trials used a selection of low redundancy, one medium and one high; half the subjects receiving them in ascending order of redundancy and half in descending order. Each subject received one selection of each type of pattern, and over the 24 all possible orders of Types A, B and C patterns were used. There were intervals between trials in which the Dvorin colour-vision test and colour matching tasks were given. No subject was noticeably colour-deficient.

The variation of type was to reduce the risks of transfer or interference which would have been likely if closely related patterns had appeared in all trials. The subjects were not told the number of different patterns appearing in any trial, so that the point at which all new patterns had appeared could not be inferred by counting.

To avoid accidental variations between replicate cards, the physically identical card appeared in repetitions. The cards were presented through a slot in a large screen on to a viewing platform. The platform could be tilted right or left with a knob, thus returning the card to the experimenter and also giving a New or Old response. No time limit was imposed, but the mechanical form of presentation tended to produce fairly prompt responses. The recording schedule distinguished:

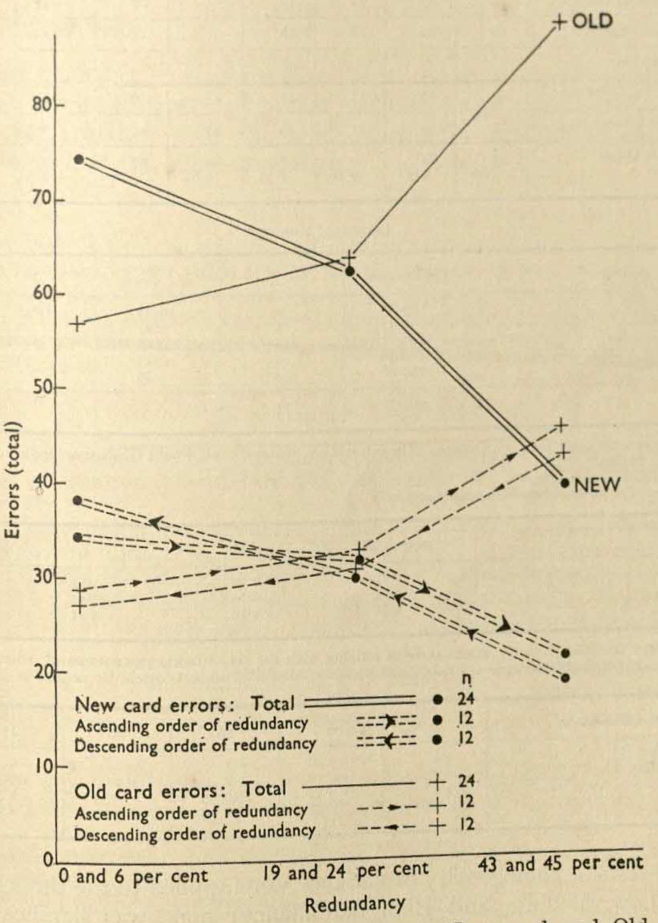
(i) "New card errors," classifying as Old a card seen for the first time.

(ii) "Old card errors," classifying as New a card which had been seen before.

*Results*

The graphs of Figure 2 summarize the results. Here the solid lines show decrease of New card errors and increase of Old card errors with increase of redundancy (Over 24 subjects). The broken-line curves below (for 12 subjects each) separate

FIGURE 2



Errors in recognition experiment, separated into New card and Old card errors. The vertical scale is of totals, for the whole group (24) in the upper curves; for the subgroups (12 each) with different orders of administration in the lower curves.

the groups receiving the sets in increasing or decreasing order of redundancy. It is obvious that the effects of order of administration are negligible compared with the effects of degree of redundancy. The symmetry of the curves about the horizontal indicates that there was little variation of the gross number of errors with redundancy or administration. It was the variation between the two types of error which offered the potentially significant result. Analysis was carried out in terms of the difference score: (New card errors—Old card errors), a positive score indicating preponderance of the first type, a negative score of the second. But, if the procedure of investigating differences between differences raises any objection, an alternative interpretation is possible and is discussed below.

TABLE II

RESULTS OF RECOGNITION EXPERIMENT

| *n* | | *New Card Errors* | | | *Old Card Errors* | | | *Differences: N − 0* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | L | M | H | L | M | H | L | M | H |
| | Ascending | | | | | | | | | |
| 6 | 0, 26, 43 per cent. | 17 | 15 | 6 | 9 | 17 | 15 | +8 | −2 | −9 |
| 6 | 6, 19, 45 per cent. | 22 | 15 | 12 | 19 | 14 | 27 | +3 | +1 | −15 |
| 12 | Totals (ascending) .. | 39 | 30 | 18 | 28 | 31 | 42 | +11 | −1 | −24 |
| | Descending | | | | | | | | | |
| 6 | 43, 26, 0 per cent. | 16 | 17 | 10 | 12 | 19 | 19 | +4 | −2 | −9 |
| 6 | 45, 19, 6 per cent. | 19 | 15 | 11 | 17 | 13 | 26 | +2 | +2 | −15 |
| 13 | Totals (descending) .. | 35 | 32 | 21 | 29 | 32 | 45 | +6 | 0 | −24 |
| 24 | Grand Totals .. | 74 | 62 | 39 | 57 | 63 | 87 | +17 | −1 | −48 |
| | Means .. .. .. | 3·08 | 2·58 | 0·16 | 2·37 | 2·62 | 3·62 | +0·70 | −0·04 | −2·0 |

*Analysis of Variance*

| *Source* | Σ *squares* | *d.f.* | *M. square* | *F* | *p* |
|---|---|---|---|---|---|
| Redundancies .. .. .. .. | 93·86 | 2 | 46·93 | 12·38 | < 0·001 |
| Orders .. .. .. .. .. | 0·22 | 1 | 0·22 | not significant | |
| Subjects* .. .. .. .. .. | 201·78 | 22 | 9·17 | 2·42 | < 0·01 |
| Interaction (orders × redundancy) .. | 0·86 | 2 | 0·43 | not significant | |
| Residual (error term) .. .. .. | 167·06 | 44 | 3·79 | | |
| Totals .. .. .. .. .. | 463·78 | 71 | | | |

* This is the error term for between Orders. The interaction, if significant, would (McNemar) indicate difference between the profiles in the blocks.

*Tests on the means of differences between conditions gave:*

| *Comparison* | *n* | *mean diff.* | *S.D.* | *t* | *p* |
|---|---|---|---|---|---|
| L: H | 24 | + 2·71 | 2·35 | 5·53 | < 0·01 |
| L: M | 24 | + 0·75 | 2·20 | 1·40 | not significant |
| M: H | 24 | + 1·96 | 4·18 | 2·24 | < 0·05 |

It may be noted that the (N − O) scores agree in ranking with the six separate percentages of redundancy, although no new significances are obtainable because the conditions *n*'s are smaller and subject correlation cannot be included:

| *Per cent. redundancy* | 0 | 6 | 19 | 26 | 43 | 45 |
|---|---|---|---|---|---|---|
| Σ scores, ascending (6) .. .. .. | +8 | +3 | +1 | −2 | −9 | −15 |
| Σ scores, descending (6) .. .. .. | +4 | +2 | +2 | −2 | −9 | −15 |
| Mean scores .. .. .. .. | +1 | +0·42 | +0·25 | −0·33 | −1·50 | −2·50 |
| S.D.s .. .. .. .. | 1·76 | 1·51 | 2·19 | 2·73 | 2·93 | 2·43 |

Bartlett tests of homogeneity of variance were applied to the three sets of results pooled as Low, Medium and High redundancy: none was significant. To take account of correlations within subjects, analysis of variance followed a model by McNemar (1962): case XV for R different and independent individuals in each block (here, orders of administration). This serves to emphasize the virtual absence of difference between orders, but otherwise shows similar F-ratios to a two-way analysis by redundancies and subjects. The results in the main groups and the significances are in Table II, and support a relationship between redundancy as assessed by this measure, and the types of error shown.

Analysing out gross errors on the different sets a trend of increasing difficulty shows from A, 115; B, 125; to C, 142, which appears as perceptually the most difficult.

It should be noted that by the administrative schedule, Old (recurrent) cards appeared three times as often as New. On a postulate of equal random guessing, errors on Old card would appear about three times as often as errors on New. The mean expectation of error by equal random guessing is 4·0 New card errors + 13·5 Old = 17·5. The mean errors per trial in the experiment was 5·58 (best subject 1·66; worst subject 13·0) so that it is clear that response is well above random level. It is a possible view that increase of Old errors represents an increase of randomness in guessing, but this is not supported by any over-all increase of errors with redundancy, i.e.:

| | Low | Medium | High |
|---|---|---|---|
| Gross error: | 131 | 125 | 126 |

Any calculation of chance levels must give the same expectation for every trial since all trials were administratively of similar pattern. And since the (N − O) score shows significant variation over the redundancies (by analysis of variance) it is highly improbable that the effect is random.

*Estimates of number*

Preliminary tests led to the selection of nine as likely to produce a suitable level of error. The average errors per subject over the three sets was 16·7, and the best subject (an artist and stage-designer) made only five errors, while only one subject completed any set without error. Each subject's responses imply an estimate of the number of different patterns involved: i.e. the number of New responses whether correct or incorrect. Thus the implied estimate is:

$$(n - \text{New errors}) + \text{Old errors}$$
$$= n - (\text{New errors} - \text{Old errors})$$

Hence these implied estimates have the same distributions as the difference score (N − O) and provide an alternative interpretation of the results, to which the same statistical significances apply.

A second estimate of number was also obtained directly by asking each subject "How many patterns do you think there were in the final set?" This question applied to the high redundancy set in the ascending trials, to the low redundancy set in the descending trials, and could be asked because the procedure gave no preliminary information on the number of patterns in any selection. The means, within fairly wide limits in both cases, were:

| | Implied estimate (from results) | Deliberate estimate (made by subject) |
|---|---|---|
| Low redundancy .. | 8·5 | 9·5 |
| High redundancy .. | 11·0 | 10·3 |

So that a possible way of stating the result is that high redundancy sets tend to be treated as having more different patterns than low redundancy sets of the same number.

Subjective reports and comments, together with one special case that will be mentioned, suggest that what may well happen is that with the high redundancy sets (in which differences are more than necessary) subjects concentrate attention on certain details and are then liable, when the same pattern recurs, to notice a detail they have not noticed before, and therefore give a New card response. The suggestion is reinforced by the behaviour of the one subject whose data were excluded from the final table. This subject gave an abnormally high number of New card responses (on one set, 22). His comments afterwards showed that he had expected trickery in the form of experiment and had decided that what he had to attend to were almost undetectable differences rather than the obvious ones. In fact, especially in the A and B sets, he concentrated on tiny variations in spacing at the junctions of colour shapes. A good deal of care had been given in preparation and these differences were seldom as much as a millimetre. Since repetitions were repetitions of the same physical card, none of these minor variations could in fact be relevant to identification: that is, two cards which were different in this respect must be clearly distinguishable in some more obvious respect, and identical cards necessarily had identical spacing. It was therefore the case that this subject had increased the redundancy of all sets by his special attitude of judgement, and with the increase of redundancy he produced an increase of Old card errors, exaggerating the tendency of more normally oriented subjects.

Yet another way of describing the difference between the low and high redundancy sets in intuitive terms, is that the low redundancy sets are more alike. Fitts (1956) makes a similar comment on his constrained set of metric figures. In these terms we might suggest that New card errors arise by confusion within the greater similarity of the low redundancy sets, and this supplies a complementary explanation of the results to that given for Old card errors with high redundancy.

In everyday speech we habitually use two kinds of judgement of likeness; we assert similarity or dissimilarity between pairs of items, and also of groups of items; e.g. "The whole family are extraordinarily alike in appearance." These two types of judgement are obviously related but not identical, the group judgement, as it were, attempting an integration of the judgements of separate pairs. One interpretation of redundancy is that it gives an integral measure of group likeness. This would open up a large prospect, but it must be remembered that the discussion here is concerned only with composite items which can be regarded as built up of distinct characteristics, not themselves analysed. The range of enquiry is thus limited. It is the working assumption of the experiment that the characteristics of colour, shape, etc., used are such as to be easily and normally discriminable from one another. The redundancy of more subtly complex figures would not be so easily referred to a plausible analysis into parts. The issue then is not merely the *order* of scanning, but the units in terms of which scanning can be stated at all.

## EXPERIMENT 2. SIMULTANEOUS DISCRIMINATION

The selections having shown variations of response in a recognition situation which called for recollection of the whole selection, it appeared relevant to ask if the results could be reduced to or related to the discriminability of the items presented as separate pairs. This was the aim of the second experiment.

*Material.* Sets B (flowers) and C (faces) were used, these having been indicated as the two more difficult sets in Experiment 1. High (43·75 per cent. and low (7·84 per cent.) redundancies only were used. The patterns were in the form of 2 in. by 2 in. transparencies.

*Procedure.* The transparencies appeared in a display apparatus (a modified Dodge tachistiscope) so arranged that any pair, whether alike or different, could be rapidly presented. The pre-exposure field was of approximately the same over-all brightness as the pattern display. The subject was instructed to respond as rapidly as possible on one of two normally-closed keys, one for "Same" and the other for "Different," so that the response was a simple disjunction. The response circuit included a selective "hold" relay which was released by a wrong response and then rejected any corrective response which followed. The response terminated the display. Response time was recorded on a Venner Electronic Stop-Clock to the nearest millisecond.

Subjects were drawn from the same sources as Experiment 1, and consisted of 10 males and 10 females. They were divided randomly into four groups which received:

(1) High redundancy (Type B) followed by low redundancy (Type C);
(2) Low ,, (Type C) ,, ,, high ,, (Type B);
(3) Low ,, (Type B) ,, ,, high ,, (Type C);
(4) High ,, (Type C) ,, ,, low ,, (Type B).

Each subject had a practice session of 72 responses on each type, followed by 108 recorded responses. The display sequence included in a random order all possible pairings of the nine patterns, balanced by an equal number of identical pairs, so that the correct responses were equalized on both keys.

*Results*

The outstanding result was the absence of any significant differences between redundancy groups, to the extent that they did not merely fall short of significance but might almost be said to prove the null hypothesis (see Table III). There was, on the other hand, a significant difference between times on B and C sets, indicating C as perceptually the more difficult and thus agreeing with the first experiment. This difference, though irrelevant to the main purpose, suggests a satisfactory form of experiment by which differences due to redundancy could have been shown if they had been of any magnitude.

TABLE III

RESULTS OF DISCRIMINATION EXPERIMENT. MEAN RESPONSE TIMES (MILLISEC.) FOR "SAME," "DIFFERENT" AND ALL RESPONSES. N = 4 × 5 = 20

| Order | *B (flower) patterns* | | | *C (face) patterns* | | |
|---|---|---|---|---|---|---|
| | *Same* | *Different* | *All* | *Same* | *Different* | *All* |
| | $B_H$ (43·75 per cent.) | | | $C_L$ (7·84 per cent.) | | |
| B → C .. .. | 736·6 | 695·9 | 716·6 | 782·2 | 698·5 | 740·5 |
| C → B .. .. | 721·0 | 653·8 | 687·0 | 904·9* | 735·3* | 822·9* |
| Means .. | 728·8 | 679·9 | 701·8 | 843·5 | 716·9 | 781·7 |
| | $B_L$ (7·84 per cent.) | | | $C_H$ (43·75 per cent.) | | |
| B → C .. .. | 728·2 | 681·4 | 710·3 | 724·2 | 704·3 | 714·8 |
| C → B .. .. | 732·5 | 674·2 | 703·0 | 834·7 | 733·9 | 785·7 |
| Means .. | 730·3 | 677·8 | 706·7 | 779·4 | 719·2 | 750·3 |

* Note: The differences are in no circumstances significant, but it may be added that the noticeable difference between $C_L$ and $C_H$ is almost entirely attributable to a single subject who was exceptionally slow on "Same" responses (1,301 millisec. compared with 805 millisec. for the other four in the same condition). Excluding this subject reduces this group's mean to 757 millisec., and emphasizes the homogeneity of the results.

In the recognition experiment, perfect performance requires memorization of each new item and the achievement of something equivalent to grasp of the structure of the set. In the second experiment, although theoretically grasp of structure might affect expectation and response, it is not essential since each discrimination can be met as a discrete event. This is compatible with the view that redundancy is not to be regarded as merely equivalent to a "sum" of differences between items, and that it is relevant when grasp of structure is involved, as, for instance, in a memory situation with that interplay of recall and generalization which can be called "coding."

## EXPERIMENT 3. DELAYED DISCRIMINATION

The third experiment was designed to explore a situation which could be considered intermediate between the previous two: a discrimination response, but with an interval between displays. Since the first pattern would have to be "committed to memory" in a sense not required in simultaneous discrimination, it was thought that the redundancy structure might again prove relevant. If so, it was predicted that the high redundancy selections, in which the variety of difference between parts was greater, would produce the longer response times. This experiment, for reasons given below, must be regarded as only a preliminary trial.

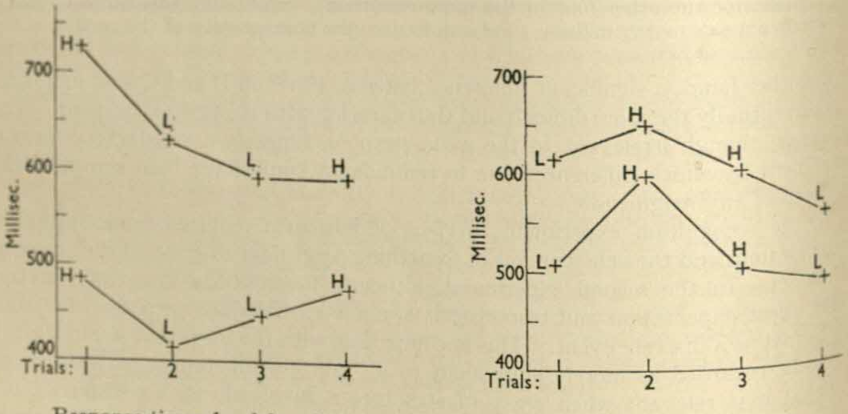
*Material.* To exclude the perceptual differences shown in previous experiments, only B (flower) sets of high (45 per cent.) and low (6 per cent.) redundancy were used. Each subject received the two selections in an abba sequence on four successive days at the same time, half starting with low and half with high redundancy. Subjects were not told whether the day's selection was the same or different from that on the previous day.

*Apparatus and procedure.* The same display apparatus was used but with further modification. The first pattern was tachistiscopically displayed for half a second without response. It appeared always in the left-hand position with the other aperture in darkness. The second pattern appeared 3·5 sec. later on the right-hand side and its exposure was terminated by the subject's response on the two-way key. At each session the subject was given 216 presentations, half again including all the possible different pairs and the other half identical pairs. The first 108 responses were treated as practice and the final 108 recorded.

### *Results*

Because of technical difficulties, the load of administration was heavy and this experiment was discontinued to give time for re-design of the apparatus and a more elaborate sequence of trials. Only four subjects (two male and two female) were used and the results cannot be offered as more than indicative. Essentially what is investigated is facilitation or delay of improvement with practice over the period of the task, and on these four subjects the results (shown as individual graphs in Fig. 3) follow the ordering of redundancies in the predicted direction, showing consistent improvement when moving from high to low redundancy selections, but tending to increase of response time when moving from low to high redundancy.

FIGURE 3



Response times for delayed discrimination. Each curve is for one subject.

It is recognized that objection can be made to the form of the experiment, but it is difficult to see any way in which first adjustments of response to selections of differing redundancy can be entirely disassociated from improvement of response by practice. All four subjects, however, were chosen as fairly highly practiced in key-response tasks, though not with this particular material or situation. Whether different degrees of redundancy retain any effect after long practice with the same material is, of course, a separate question.

## COMMENTS

(1) The primary theoretical attempt was to give the measure of redundancy a legitimately generalized form which could be applied to visual aggregates irrespective of any assumption of scanning order. The internal consistency of the results in Experiment 1, with variations of response related to redundancy dominating over the perceptual differences between the three types of pattern, suggests that a genuine variable has been obtained, though its range of application can only be determined by an extended series of experiments.

(2) An advantage of the measure is that it is calculated directly from the determinates appearing in the experimental selection and does not involve any preliminary decision as to whether the selection has been taken "randomly" or "non-randomly" from a theoretical source. Selections which would equally be random by method of selection can thus show different levels of redundancy, and variations of response with at least three levels of redundancy were obtained, with some suggestion of variation over six levels. It might be added that repetitions of the recall experiments with Miller's random and redundant strings (Staniland, 1960) have reinforced the results then reported, and that efficiency of recall relates even more closely to D-redundancy than to the E-redundancy measure first used.

(3) Redundancy is essentially a characteristic of a selection taken as a whole, not a property attributable to items taken in isolation. The conclusion suggested by these experiments is in accordance with this fact: that redundancy may be important when the nature of the task requires grasp of the over-all structure of display, but less relevant when (as in simultaneous discrimination) immediate response can be made without reference to over-all structure. Whether in the relatively long term recall required for recognition, or the shorter-term recall for delayed discrimination, what is suggested is that the low-redundancy selections allow easier coding for storage than high-redundancy selections with a superfluity of differences.

This paper is extracted from a thesis accepted by the University of Exeter for the degree of Ph.D. The experimental work was helped by a grant from the University. Thanks are due to the supervisor, Dr. J. Szafran, for discussion and criticism.

## APPENDIX

*Mathematical Note*

The value $T_{1\,..\,c}$ which is the numerator in the D-redundancy measure is the same as Garner's sum of contingent uncertainties. For:

$$\begin{aligned} T_{1:2} &= T_{1:2} \\ T_{1,2:3} &= T_{1:3} + T_{2:3} + A_{123} \\ T_{1,2,3:4} &= T_{1:4} + T_{2:4} + T_{3:4} + A_{124} + A_{134} + A_{234} + A_{1234} \\ \text{so that: } \Sigma T_{1\,...\,4} &= T_{1:2} + T_{1:3} + T_{1:4} + T_{2:3} + T_{2:4} + T_{3:4} + A_{123} + A_{124} + \\ &\qquad A_{134} + A_{234} + A_{1234} \\ &= \text{Garner's sum of contingent uncertainties.} \end{aligned}$$

Garner (1962) says (p. 154) that:

"In order to retain consistency of meaning of the total constraint, it would be better not to include distributional constraint. This problem is very easily avoided by computing the total constraint with . . . the actual uncertainties of all variables rather than with the nominal uncertainties."

The denominator of D-redundancy does precisely this, by taking $\Sigma - p \log_2 p$ for each determinable, instead of setting each at the nominal value of $\log_2 s$, as in Shannon's denominator, $c \log_2 s$.

## REFERENCES

ATTNEAVE, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.*, **61**, 183–93.

BRUNER, J. S., GOODNOW, J., and AUSTIN, G. A. (1956). *A Study of Thinking.* New York.

CHERRY, C. (Ed.) (1961). *Information Theory. Fourth London Symposium.* London.

FITTS, P. M., WEINSTEIN, M., RAPPAPORT, M., ANDERSON, N., and LEONARD, J. A. (1956). Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, **51**, 1–11.

GARNER, W. R. (1962). *Uncertainty and Structure as Psychological Concepts.* New York.

JOHNSON, W. E. (1921). *Logic.* Cambridge.

MCNEMAR, Q. (1962). *Psychological Statistics.* New York.

MILLER, G. A. (1958). Free recall of redundant strings of letters. *J. exp. Psychol.*, **56**, 485–91.

SHANNON, C. E., and WEAVER, W. (1949). *The Mathematical Theory of Communication.* Illinois.

STANILAND, A. (1960). Redundancy as an experimental variable. *Quart. J. exp. Psychol.*, **12**, 149–61.

*Manuscript received 8th October,* 1962.

# THE PERSISTENCE OF SOME EFFECTS OF HANDLING IN INFANCY ON THE BEHAVIOUR OF THE ADULT RAT

BY

P. D. DU PREEZ

*From the Department of Psychology, University of Cape Town*

An attempt was made to discover how persistent the effects of handling before weaning are on activity, grooming, emotional elimination and learning in adult rats; whether there is an interaction of the effects of handling and the effects of cage size on the behaviour of rats; and whether the daily removal of the mother during nursing has any effects on the pups.

Since all differences between handled and undisturbed animals found at 50 or 100 days were found at 220 days, it was concluded that the effects of handling in infancy are persistent. The hypothesis that cage size and removal of the mother during nursing significantly affect adult behaviour was not supported by the results.

## INTRODUCTION

Handling has been found to raise the resistance of rats to immobilization (Weininger, 1953), to glucose injection (Levine, 1957*a*), and to cold (Levine, Alpert and Lewis, 1958). The growth rate of handled animals is greater, food utilization is better, and thyroid activity is lower (Ruegamer, Bernstein and Benjamin, 1954).

Behavioural characteristics of handled rats are tameness and greater ambulation in the open field (Weininger, 1956), less emotional behaviour, such as grooming, freezing and defecation in the test situation (Levine, 1956), more rapid consummation after water deprivation (Levine, 1957*b*), greater dominance (Mezei and Rosen, 1960), and more rapid acquisition of an avoidance response (Levine, Chevalier and Korchin, 1955).

The question of the persistence of these effects does not seem to have been raised. A review of the ages at which animals were tested for the effects of handling shows that the most striking differences were found with animals tested relatively young. Weininger (1953, 1956), commenced testing his animals at nearly 60 days, Levine (1956, 1957*a*, 1957*b*) commenced with his oldest animals between the ages of 60 and 70 days, Ruegamer *et al.* (1954) tested their animals at 56 days, and Mezei and Rosen (1960) tested their animals at 47 days. These experimenters found significant differences between their handled and control groups, but it is not known whether these differences would have persisted.

The following experimenters, who did not find pronounced differences caused by handling, tested their animals at rather later ages. Gertz (1957) tested animals at 120 days which had had only one experience of handling; Scott (1955) commenced testing his animals at 119 days, but reduced the value of his experiment by handling the control group once a week for weighing; and Ader (1959) tested his late-handled group at 181 days, only 15 days after their last experience of handling. These studies do not provide conclusive evidence for any lack of permanence in the effects of handling, and it remains to be shown that the effects of early handling can continue to affect the behaviour of adult animals for long periods.

The first problem to which the present experiment is directed is the persistence of the effects of handling before weaning on growth rate, activity, grooming, emotional elimination, avoidance learning and elevated maze learning.

Another possible source of discrepancy in studies of the effects of early handling is the size and nature of the cage. The cages used by Ader (1959) allowed a free view of the room, while the cages employed by Levine (1956) were somewhat restrictive, preventing the animals from having much of a view of their environment. Cages used by Ader were also larger than those used by Levine.

It was decided to investigate the interaction of the effects of handling and the effects of cage size on behaviour because cage size may be an important stimulus variable in the life of laboratory animals which may serve to either increase or decrease the effects of handling.

It was also decided to investigate the effects of daily removal of the mother for the equivalent of the handling period of 15 min. This was done to discriminate between the effects of handling and the effects of removal of the mother on the behaviour of the handled group.

## DESIGN

A total of 126 subjects, representing five types of treatment, was divided into three groups, each group to be tested at a different age.

The five treatments were:

(i) sh: handled and housed in small cages (8 in. × 12 in. × 6 in.);
(ii) lh: handled and housed in large cages (57 in. × 10 in. × 10 in.);
(iii) su: undisturbed, housed in small cages;
(iv) lu: undisturbed, housed in large cages;
(v) sm: mother removed daily for equivalent of handling period of 15 min.; but untouched by hand and housed in small cages.

Distribution of subjects is given in table below.

TABLE I

DISTRIBUTION OF SUBJECTS

| | *Subjects* | | *Treatment* | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Group* | *Age at test* | *Sex* | *sh* | *lh* | *su* | *lu* | *sm* | *Total* |
| G50 .. | 50 days | male | 10 | 5 | 10 | 5 | 10 | 40 |
| G100 .. | 100 days | male | 4 | 4 | 4 | 4 | 4 | 20 |
| | | female | 8 | 4 | 8 | 4 | 8 | 32 |
| G220 .. | 220 days | female | 8 | 5 | 8 | 5 | 8 | 34 |
| Total .. | | | 30 | 18 | 30 | 18 | 30 | 126 |

## METHOD

### *Apparatus*

Avoidance training apparatus similar to that described by Miller (1948) and an elevated T-maze, were used.

The avoidance training apparatus was a glass-walled box measuring 24 in. × 10 in. × 12 in., divided into two equal compartments by a perspex wall in which was set, at a height of 5 in., an aperture 3 in. × 3 in. The aperture was kept open during preliminary habituation trials and closed by a one-way door during avoidance learning trials. Illumination was by a 40-watt bulb at 6 ft. The grid floor was electrified by a 12 volt current from two dry cells through an induction coil. A buzzer was connected in the circuit.

The elevated T-maze consisted of two 3-ft. pieces 30 in. high and 1 in. across at the top. Illumination was by a 100-watt bulb at 10 ft.

*Subjects*

The females selected for breeding were of the same age, 110 days, and were of the Wistar strain kept in the Department of Physiology of the University of Cape Town. During the gestation period they were handled in order to tame them so that they could be removed from their cages without difficulty when the time came for handling the pups. Immediately they had given birth, the number of pups per litter was brought down to six to eliminate, as far as possible, differences in nutrition and competition.

Pregnant females were housed individually. Entire litters were randomly assigned to each treatment (handling, ignoring, or removing the mother for 15 min. each day).

*Procedure*

(i) *Treatment*. Handling occurred daily and followed the routine given below.

After the mother had been removed from her litter and placed in another cage, the pups were individually picked up between the thumb and the index finger of the left hand and transferred to the palm of the right hand. Each pup was stroked along its back from the base of the skull to the tail with the index finger of the left hand at a rate of approximately one stroke per sec. After 30 sec. the pup was transferred to a small compartment (3 in. × 3 in.) in a cardboard box. After 9 min. the pup was returned to its nest. The mother was returned 15 min. after removal.

The ignored group never had the mother removed, and the mother-removed group had the mother removed for 15 min., but the pups were not handled.

Handling commenced not more than 24 hr. after birth. When a new litter assigned to handling treatment was found during the daily inspection of cages, handling of the pups was commenced immediately.

At 25 days the pups were weaned, weighed, sexed and marked on their ears before random assignment to large or small cages. Both the large and the small cages had chicken wire front and top and solid back and sides. All animals were able to look freely through the cage front. Droppings and waste fell through the grid into trays of sawdust which could be removed and cleaned without handling the animals.

Four animals were housed in each cage in a room kept at a temperature of 70° F. They were not handled between weaning and testing.

Food and water were available ad lib.

Subjects were in three groups, each containing all five types of treatment, to be tested at three different ages: 50 days, 100 days, and 220 days.

Thirteen litters supplied the animals of G50 and G220. Of these, 37 were females and 41 were males; but three females and one male died after weaning. The males were allotted to G50 and the females to G220, since it was decided to keep groups unisexual in order to reduce standard deviation which might obscure differences caused by treatment. It was not supposed that there would be a significant interaction between sex and treatment, but it was thought that mean scores for the sexes might be different. To check this, nine more litters, composed of 22 males and 32 females, were bred later to be tested at 100 days. Two of the males of this group and none of the females died after weaning.

The number of animals in each cage was brought up to four, in some cases by using additional animals from litters bred at the same time which did not fulfil all requirements for inclusion in the experimental groups (litter less than six, members of the litter died before weaning).

(ii) *Testing*. All groups were tested in the same way. G50 commenced testing at 50 days, G100 commenced testing at 100 days, and G220 commenced testing at 220 days. Each animal was weighed on the first day of testing. Each series of tests lasted from 3 to 4 weeks.

(A) Each animal was given three preliminary runs of 3 min. each, at the rate of one per day in the avoidance training apparatus. No shock was given in these trials. Counts were taken of rears, crossing between compartments, latency to first crossing in each 3-min. run, grooms and elimination.

(B) Avoidance training was commenced at the rate of two spaced trials (90-min. interval) per day and was continued until the animal had reached a criterion of two successive avoidances of shock by escape into the adjacent compartment. Each trial was conducted as follows: The animal was placed in the starting compartment, a buzzer was sounded for 5 sec., and the shock commenced 5 sec. after the buzzer had stopped sounding. If at the end of 50 sec. of shock the animal had not escaped, it was assisted through the aperture. Each animal was returned to its home cage immediately it had completed the avoidance response. The number of trials to establish the avoidance response and the latency to response were recorded.

(C) Extinction trials commenced immediately a subject had learnt the avoidance response and continued until a criterion of two successive no-responses of 30 sec. each had been reached. Two spaced trials (90-min. interval) were run per day. The number of trials to extinguish the avoidance response and the latency to response were recorded.

(D) Subjects learnt the elevated T-maze at the rate of one trial per day for 5 days for a water reward under conditions of 21 hr. water deprivation. Each animal was allowed to remain on the maze until he had reached the goal. The situation was evidently fear-provoking and differences in emotionality should have shown up. The reward was 10 sec. consumption of water. Time to reach goal and time to consume water were noted. Levine (1957*b*) found that handled animals consumed more rapidly after water deprivation.

## RESULTS

### *Comparison of differences between handled and undisturbed animals*

(1) *Rate of growth* as revealed by the ratio of weight at testing age to weight at 25 days was not affected by any of the treatments.

(2) *Activity* was measured by rears, crossings between compartments, and latency to first crossing in the preliminary trials.

At all three ages the handled animals were found to rear more than the undisturbed animals at better than the 1 per cent. level of confidence.

TABLE II

F VALUES FOR DIFFERENCES BETWEEN HANDLED AND UNDISTURBED ANIMALS

| Tests | *Groups* | | | |
|---|---|---|---|---|
| | *G50 days male* | *G100 days* | | *G220 days female* |
| | | *male* | *female* | |
| A. *Preliminary* | | | | |
| Rears.. .. .. .. | 18·2** | 20·1** | 6·3* | 11·4** |
| Crossings .. .. .. | 1·6 | 14·7** | 1·2 | 21·7** |
| Latency .. .. .. | 1·1 | 13·9** | 2·8 | 17·4** |
| Grooms .. .. .. | 5·4* | 0·3 | 0·3 | 4·75* |
| B. *Avoidance learning* | | | | |
| Trials.. .. .. .. | 4·26* | 0·2 | 0·1 | 8·1** |
| Latency .. .. .. | 3·8 | 0·0 | 0·6 | 8·03** |
| C. *Extinction* | | | | |
| Trials.. .. .. .. | 0·0 | 0·1 | 2·4 | 5·2* |
| Latency .. .. .. | 1·6 | 0·1 | 1·9 | 1·5 |
| D. *Elevated maze* | | | | |
| Time to goal .. .. | 2·8 | 0·2 | 0·8 | 2·53 |
| Time to first consumption .. | 1·6 | 0·2 | 1·2 | 2·4 |
| Errors .. .. .. | 0·3 | 0·1 | 0·4 | 1·0 |
| E. *Weight increase* .. .. | 1·2 | 0·3 | 0·1 | 2·3 |

* = 5 per cent. level of confidence
** = 1 per cent. level of confidence

Handled males crossed the barrier more frequently at 100 days than undisturbed males, and handled females crossed the barrier more frequently than undisturbed females at 220 days. No differences were found between handled and undisturbed males at 50 days or females at 100 days. The combined F value for males and females at 100 days ($F = 3{\cdot}9$, $p < 0{\cdot}05$) reveals significant differences between handled and undisturbed animals.

Latency to first crossing revealed differences between handled and undisturbed animals in the same classes as crossing scores above. Differences failed to appear among the males at 50 days and the females at 100 days. A combined male–female F value at 100 days revealed that handled subjects had a smaller latency to first crossing than undisturbed subjects at better than the 1 per cent. level of confidence ($F = 8{\cdot}96$).

No measure of activity was affected by either cage size or removal of the mother. Interaction scores between handling and cage size were non-significant.

(3) *Grooms.* Undisturbed animals groomed more at 50 days and at 220 days. No differences appeared at 100 days. Grooming of animals at 50 days and at 100 days was significantly increased by housing in large cages as compared to housing in small cages. Interaction scores of some significance were found at 50 days ($F = 4{\cdot}53$, $p < 0{\cdot}05$) and at 220 days ($F = 6{\cdot}0$, $p < 0{\cdot}05$). Removal of the mother did not significantly affect scores.

(4) *Elimination.* Under the conditions of our experiment elimination was found to be restricted to a small number of individuals scattered eccentrically among the treatments. Emotional elimination in the testing apparatus did not distinguish among treatment types.

(5) *Avoidance learning.* G50 and G220 handled animals learnt the avoidance response in fewer trials than the undisturbed animals. No differences were found in G100. Only in G220 did the handled animals have a smaller mean response latency for all trials for each rat than the undisturbed animals. Cage size produced no significant effects. Neither did removal of the mother.

(6) *Extinction trials.* Only at 220 days did handled animals extinguish the avoidance response faster than undisturbed animals. Neither cage size nor removal of the mother significantly affected results.

(7) *Elevated maze.* Handling, cage size, and removal of the mother did not affect scores on the elevated maze.

(8) *Sex differences* in performance were analysed in group 100. They were found in rears ($F = 6{\cdot}8$, $p < 0{\cdot}05$), crossings ($F = 7{\cdot}1$, $p < 0{\cdot}05$). Females were more active and required more trials to extinguish the avoidance response. They groomed less than males did. No significant interaction values between sex and handling were found, so that we can safely assume that, though the level of scoring for males and females is different, the reaction to the experimental treatment is the same.

### *Comparison of average scores at three different ages*

(1) *Stable scores.* The level of rears, grooms, trials to learn the avoidance response, trials to extinguish the avoidance response, latency in making avoidance response during extinction trials, and errors on the elevated maze is almost identical at all ages.

(2) *Scores showing increase in activity.* The level of latency to first crossing in the preliminary trials tends to decrease slightly with age, which means that older

completed the avoidance response. The number of trials to establish the avoidance response and the latency to response were recorded.

(C) Extinction trials commenced immediately a subject had learnt the avoidance response and continued until a criterion of two successive no-responses of 30 sec. each had been reached. Two spaced trials (90-min. interval) were run per day. The number of trials to extinguish the avoidance response and the latency to response were recorded.

(D) Subjects learnt the elevated T-maze at the rate of one trial per day for 5 days for a water reward under conditions of 21 hr. water deprivation. Each animal was allowed to remain on the maze until he had reached the goal. The situation was evidently fear-provoking and differences in emotionality should have shown up. The reward was 10 sec. consumption of water. Time to reach goal and time to consume water were noted. Levine (1957*b*) found that handled animals consumed more rapidly after water deprivation.

## RESULTS

### *Comparison of differences between handled and undisturbed animals*

(1) *Rate of growth* as revealed by the ratio of weight at testing age to weight at 25 days was not affected by any of the treatments.

(2) *Activity* was measured by rears, crossings between compartments, and latency to first crossing in the preliminary trials.

At all three ages the handled animals were found to rear more than the undisturbed animals at better than the 1 per cent. level of confidence.

TABLE II

F VALUES FOR DIFFERENCES BETWEEN HANDLED AND UNDISTURBED ANIMALS

| Tests | *Groups* | | | |
|---|---|---|---|---|
| | *G50 days male* | *G100 days* | | *G220 days female* |
| | | *male* | *female* | |
| A. *Preliminary* | | | | |
| Rears | 18·2** | 20·1** | 6·3* | 11·4** |
| Crossings | 1·6 | 14·7** | 1·2 | 21·7** |
| Latency | 1·1 | 13·9** | 2·8 | 17·4** |
| Grooms | 5·4* | 0·3 | 0·3 | 4·75* |
| B. *Avoidance learning* | | | | |
| Trials | 4·26* | 0·2 | 0·1 | 8·1** |
| Latency | 3·8 | 0·0 | 0·6 | 8·03** |
| C. *Extinction* | | | | |
| Trials | 0·0 | 0·1 | 2·4 | 5·2* |
| Latency | 1·6 | 0·1 | 1·9 | 1·5 |
| D. *Elevated maze* | | | | |
| Time to goal | 2·8 | 0·2 | 0·8 | 2·53 |
| Time to first consumption | 1·6 | 0·2 | 1·2 | 2·4 |
| Errors | 0·3 | 0·1 | 0·4 | 1·0 |
| E. *Weight increase* | 1·2 | 0·3 | 0·1 | 2·3 |

* = 5 per cent. level of confidence
** = 1 per cent. level of confidence

Handled males crossed the barrier more frequently at 100 days than undisturbed males, and handled females crossed the barrier more frequently than undisturbed females at 220 days. No differences were found between handled and undisturbed males at 50 days or females at 100 days. The combined F value for males and females at 100 days ($F = 3{\cdot}9$, $p < 0{\cdot}05$) reveals significant differences between handled and undisturbed animals.

Latency to first crossing revealed differences between handled and undisturbed animals in the same classes as crossing scores above. Differences failed to appear among the males at 50 days and the females at 100 days. A combined male–female F value at 100 days revealed that handled subjects had a smaller latency to first crossing than undisturbed subjects at better than the 1 per cent. level of confidence ($F = 8{\cdot}96$).

No measure of activity was affected by either cage size or removal of the mother. Interaction scores between handling and cage size were non-significant.

(3) *Grooms.* Undisturbed animals groomed more at 50 days and at 220 days. No differences appeared at 100 days. Grooming of animals at 50 days and at 100 days was significantly increased by housing in large cages as compared to housing in small cages. Interaction scores of some significance were found at 50 days ($F = 4{\cdot}53$, $p < 0{\cdot}05$) and at 220 days ($F = 6{\cdot}0$, $p < 0{\cdot}05$). Removal of the mother did not significantly affect scores.

(4) *Elimination.* Under the conditions of our experiment elimination was found to be restricted to a small number of individuals scattered eccentrically among the treatments. Emotional elimination in the testing apparatus did not distinguish among treatment types.

(5) *Avoidance learning.* G50 and G220 handled animals learnt the avoidance response in fewer trials than the undisturbed animals. No differences were found in G100. Only in G220 did the handled animals have a smaller mean response latency for all trials for each rat than the undisturbed animals. Cage size produced no significant effects. Neither did removal of the mother.

(6) *Extinction trials.* Only at 220 days did handled animals extinguish the avoidance response faster than undisturbed animals. Neither cage size nor removal of the mother significantly affected results.

(7) *Elevated maze.* Handling, cage size, and removal of the mother did not affect scores on the elevated maze.

(8) *Sex differences* in performance were analysed in group 100. They were found in rears ($F = 6{\cdot}8$, $p < 0{\cdot}05$), crossings ($F = 7{\cdot}1$, $p < 0{\cdot}05$). Females were more active and required more trials to extinguish the avoidance response. They groomed less than males did. No significant interaction values between sex and handling were found, so that we can safely assume that, though the level of scoring for males and females is different, the reaction to the experimental treatment is the same.

### *Comparison of average scores at three different ages*

(1) *Stable scores.* The level of rears, grooms, trials to learn the avoidance response, trials to extinguish the avoidance response, latency in making avoidance response during extinction trials, and errors on the elevated maze is almost identical at all ages.

(2) *Scores showing increase in activity.* The level of latency to first crossing in the preliminary trials tends to decrease slightly with age, which means that older

### TABLE III

### Average Scores of Handled and Undisturbed Animals

| | *Groups* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *G50 days* | | *G100 days* | | | | *G220 days* | |
| | *male* | | *male* | | *female* | | *female* | |
| *Tests* | *h* | *u* | *h* | *u* | *h* | *u* | *h* | *u* |
| A. *Preliminary* | | | | | | | | |
| Rears .. .. | 72 | 57 | 70 | 51 | 82 | 64 | 84 | 60 |
| Crossings .. .. | 7·3 | 4·7 | 12·4 | 6·3 | 16·3 | 13·3 | 13·8 | 4·8 |
| Latency .. .. | 361 | 418 | 193 | 356 | 176 | 266 | 159 | 369 |
| Grooms .. .. | 1·4 | 2·7 | 2·6 | 3·1 | 1·2 | 1·3 | 1·2 | 2·9 |
| B. *Avoidance learning* | | | | | | | | |
| Trials .. .. | 4·0 | 5·4 | 4·1 | 3·9 | 4·0 | 4·3 | 4·5 | 8·7 |
| Latency .. .. | 103·7 | 184·6 | 58 | 57·8 | 41·4 | 59 | 84·1 | 356 |
| C. *Extinction* .. | | | | | | | | |
| Trials .. .. | 10·6 | 10·7 | 15·1 | 15·7 | 17·6 | 23·8 | 17·0 | 24·6 |
| Latency .. .. | 83·8 | 86·3 | 100 | 105 | 126 | 93 | 120 | 94 |
| D. *Elevated maze* | | | | | | | | |
| Total time to goal | 163·7 | 230·8 | 340 | 363 | 420 | 483 | 466 | 608 |
| Total time to first consumption .. | 117·0 | 163·5 | 266 | 354 | 354 | 421 | 438 | 587 |
| Total errors .. | 11 | 13 | 10 | 11 | 12 | 12·7 | 13·5 | 17·2 |
| E. *Weight increase* .. | 33·1 | 34·3 | 68·8 | 64·5 | 48·5 | 50·5 | 53·5 | 57·5 |

*h* = handled
*u* = undisturbed

animals tend to cross the barrier from the starting compartment into the adjacent rather sooner than younger animals. Frequency of crossing is also higher at 220 days and 100 days than at 50 days.

(3) *Scores showing decrease in activity.* On the elevated maze the total time taken to reach goal and the total time to consume water for the first time tend to increase with age.

## Discussion

### (i) *Handling*

Whatever differences appear at 50 days also appear at 220 days. The first conclusion we are able to draw is that such differences as there are between handled and undisturbed animals are persistent.

Does the evidence also permit us to conclude that there is a trend towards increase or diminution of the differences? The striking thing is that differences at 220 days tend to be greater than differences at 100 days or at 50 days. Differences in activity scores between handled and undisturbed animals are greater at 100 days than at 50 days, but differences in grooming and avoidance learning are less. The problem is whether we are dealing with two distinct trends. Do activity differences due to handling show a uniform increase with age, and do grooming and avoidance learning differences show a decline at 100 days? That is what our evidence indicates, but the explanation is not clear.

The possibility that large overriding changes in the actual level of performance might be responsible for these two tendencies is not strongly supported by an examination of average scores. There is no clear tendency for scores at 100 days to be either higher or lower than scores at other ages.

How can we explain the increase in differences between handled and undisturbed animals with age? It appears that two types of explanation could be offered, both starting with the assumption that a higher level of fear causes a lower level of activity (Montgomery, 1955).

The first and most obvious explanation is that early-handled animals are tamer and respond with less fear to handling in the tests. Previously undisturbed animals might reasonably be assumed to tame more easily at 50 days than at 100 days or 220 days. The conclusion from this would be that the previously undisturbed animals would adjust more rapidly the earlier the testing and show less difference from the early-handled group.

If this explanation is correct, then we should expect the first tests in the series to show greater differences between the handled and undisturbed animals than the later tests. This is so. More differences appear in the preliminary runs than in the avoidance, extinction, or elevated maze trials. A second though not necessary deduction, that the first preliminary trial should distinguish more definitely between handled and undisturbed animals, is not confirmed by an analysis of the results. Nor is the possibility that these groups differ rather in their rate of habituation in the preliminary runs than in first reaction to the extreme novelty of the first run supported by any consistent trend in their scores. The data of G50, G100 and G220 show no single pattern of increasing or decreasing difference between handled and non-handled groups. Nor is there a consistent pattern within any age group which would enable one to draw conclusions about the rôle of age.

A further source of evidence is the actual level of performance at the three ages. If the differences are greater in later tests because animals are more difficult to tame at a later than at an earlier age, we might expect that the undisturbed group would be less active at 220 days than at 100 days, or at 100 days than at 50 days. The tendency is not very definite, so that again we can draw no firm conclusions. It is possible that handling has other effects in addition to the taming effect.

It is this second possibility which has received most attention. It has been hypothesized by various writers (Ader, 1959; Levine, 1960) that early stimulation—of which handling is an example—produces changes in adrenal responsiveness. It is known that early handling accelerates the appearance of the adrenal response to cold (Levine *et al.*, 1958) and decreases the total adrenal response to immobilization (Weininger, 1956). There is evidence that handled animals respond to stress by a more rapid rise in the level of adrenalin in the blood, but that the level also declines more rapidly (Levine, 1960).

Deductions from this last observation would be that handled animals should acquire the avoidance response more rapidly and extinguish more rapidly. This is partially borne out by the fact that handled animals require fewer trials at 50 and 220 days to learn avoidance. But only at 220 days do they extinguish in fewer trials. The importance of adrenal secretion in maintaining an avoidance response has been shown with dogs (Wynne and Solomon, 1955). Adrenalectomized animals lost the avoidance response within one or two trials whereas intact animals continued to make avoidance responses for hundreds of trials.

The exact way in which adrenal responsiveness of early-handled animals is altered and the relationship of this to the entire functioning of the nervous system has not yet been clarified. And it is not clear yet precisely what kinds of early stimulation will produce such an alteration, and what relationship it has to taming.

The results cannot be accounted for entirely by habituation to contact with the experimenter. If that were so, differences would be at their peak at 50 days, and then decline as a function of the time elapsed since the last experience of handling. But the additional assumption that it becomes more difficult to tame an animal the older it is before taming is begun, remains a possible explanation of the apparent increase of some differences between handled and undisturbed animals as they grow older.

### (ii) *Cage size*

Animals from the large and small cages showed few differences. Since both sets had a free view of the room, the difference in their experience was probably proprioceptive. One might have expected differences in level of activity, but this expectation was not confirmed. That there were no differences in learning speed on the elevated maze confirms Forgus' (1954) finding that there is little difference between the maze learning ability of a "complex visual and complex proprioceptive group and a complex visual and poor proprioceptive group." It is possible that visual restriction would produce greater interaction with handling.

### (iii) *Removal of the mother*

The removal of the mother for 15 min. a day did not affect behaviour variables in this experiment. Possibly, if the mother had not been carefully tamed before the experiment, different results might have been obtained. The mother's reaction to removal may be more important than the pups'.

Untamed mothers are extremely agitated by removal from their nests. When they are returned they usually remove the entire litter to another part of the cage, tread on the young, and lick them. If the mothers are not previously tamed, removal may still be an important variable to control in experiments of this kind.

## REFERENCES

ADER, R. (1959). The effects of early experience on subsequent emotionality and resistance to stress. *Psychol. Monogr.*, **73**, 1–31.

FORGUS, R. H. (1954). The effects of early perceptual learning on the behavioural organization of adult rats. *J. comp. physiol. Psychol.*, **47**, 331–6.

GERTZ, B. (1957). The effect of handling at various age levels on emotional behaviour of adult rats. *J. comp. physiol. Psychol.*, **50**, 613–6.

LEVINE, S. (1956). A further study of infantile handling and adult avoidance learning. *J. Person.*, **25**, 70–80.

LEVINE, S. (1957*a*). Infantile stress and resistance to physiological stress. *Science*, **126**, 3270.

LEVINE, S. (1957*b*). Infantile stress and consummatory behaviour in adulthood. *J. comp. physiol. Psychol.*, **50**, 609–12.

LEVINE, S. (1960). Stimulation in infancy. *Scientific American*, **202**, 80–101.

LEVINE, S., ALPERT, M., and LEWIS, G. W. (1958). Differential maturation of an adrenal response to cold stress in rats handled in infancy. *J. comp. physiol. Psychol.*, **51**, 774–7.

LEVINE, S., CHEVALIER, J. A., and KORCHIN, S. J. (1955). The effects of shock and handling in infancy on later avoidance learning. *J. Person.*, **24**, 475–93.

MEZEI, T., and ROSEN, J. (1960). Dominance behaviour as a function of infantile stimulation in the rat. *Arch. gen. Psychol.*, **3**, 53–6.

MILLER, N. E. (1948). Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, **38**, 89–101.

MONTGOMERY, R. (1955). The relation between fear and exploratory behaviour. *J. comp. physiol. Psychol.*, **48**, 254–60.

RUEGAMER, W. R., BERNSTEIN, L., and BENJAMIN, J. D. (1954). Growth, food utilization, and thyroid activity in the albino rat as a function of extra handling. *Science*, **120**, 184–5.

SCOTT, J. H. (1955). Some effects at maturity of gentling, ignoring, or shocking rats during infancy. *J. abnorm. soc. Psychol.*, **51**, 412–4.

WEININGER, O. (1953). Mortality of albino rats under stress as a function of early handling. *Canad. J. Psychol.*, **7**, 111–4.

WEININGER, O. (1956). The effects of early experience on behaviour and growth characteristics. *J. comp. physiol. Psychol.*, **49**, 1–9.

WYNNE, L. C., and SOLOMON, R. L. (1955). Traumatic avoidance learning: Acquisition and extinction in dogs deprived of peripheral autonomic function. *Genet. Psychol. Monogr.*, **52**, 241–84.

*Manuscript received* 8*th October*, 1962.

# TACTILE DISCRIMINATION OF SHAPE BY OCTOPUS

BY

M. J. WELLS

*From the Department of Zoology, Cambridge*

Experiments are described showing that *Octopus* can be trained to distinguish by touch between a cube and a sphere, a surprising result since previous work has always indicated that octopuses cannot take proprioceptive information into account in learning. It can be shown, however, that the cube/sphere discrimination is made by detecting the distortion imposed on the rims of individual suckers in contact with the corners of the cube. Thus it is shown (1) that a cube with the corners even slightly rounded is less readily distinguishable from a sphere than a cube with sharp corners and, more important, (2) that a narrow rod is readily accepted in place of the cube by trained animals. Indeed, the rod seems to be a "better" cube than the original, and errors of discrimination decrease when it replaces the cube, presumably because it lacks flat surfaces liable to be confused with the surface of the sphere by a system depending on information from sense organs arranged in circles on the rims of the suckers.

## INTRODUCTION

Octopuses can readily be taught to distinguish between objects by touch, provided that they differ in texture. Objects differing in weight, or in the orientation or pattern of surface irregularities, are not distinguished. It has been argued that these results show octopuses to be incapable of taking proprioceptive information into account in tactile learning, for the discriminations that they fail to make are those that would require the integration of information from internal and superficial receptors (Wells, 1961*a*, 1962; Wells and Wells, 1957).

Shape discrimination, which can be done only by making use of information from internal receptors to locate the relative positions of the surface sense organs registering the contact, should therefore be impossible for octopuses. The experiments reported here were made to check this.

## MATERIAL

*Octopus vulgaris* Lamarck of from 250 to 500 gm. from the Bay of Naples were used. The animals were blinded by section of the optic nerves (Wells and Wells, 1956). Brain lesions to the inferior frontal system (Wells, 1959) were checked from serial sections.

## METHOD

The animals were trained using a reward and punishment technique that has been fully described elsewhere (Wells and Wells, 1956). Briefly, each experiment consisted of a number of trials at each of which one or other of a pair of objects was presented to the octopus by touching it against the back of one of the arms. The animal would then twist the arm to grasp the object with the suckers, bend the arm and pass it under the interbrachial web to the mouth. The octopus was rewarded with a piece of sardine for taking the "positive" object in this way, and punished by means of a 6–9 volt A.C. shock for taking the "negative," with the result that after the first few trials a proportion of the test objects were rejected, by thrusting them away instead of passing them in towards the mouth. In the first part of the experiment to be reported, there were eight trials per day at intervals of not less than one hour, arranged thus: + − + − + + − −, next day − + − + + − − +, and so on. Later, the rate of training was raised to 40 trials per day, in two groups of 20, the trials of each group in the following sequence: + − + − + + − − + − + − + − − + + − + −.

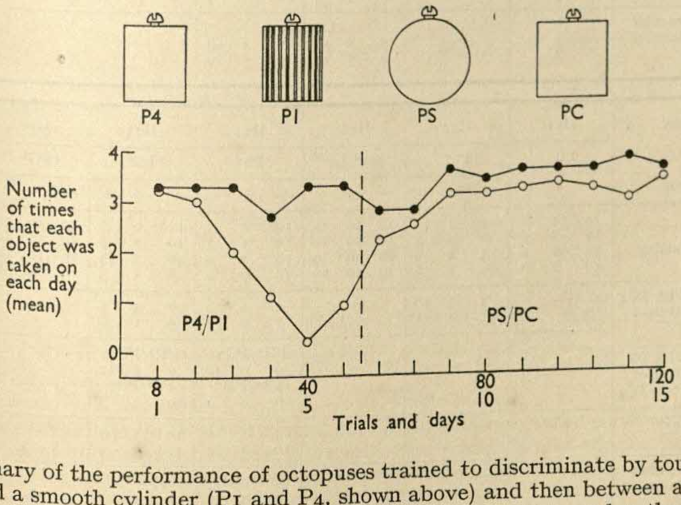
A large number of experiments have been made under these conditions and there is no evidence to suggest that octopuses can learn to recognize the trials' sequence (Wells, 1962). The trials were arranged in these sequences rather than at random because *Octopus* learns so rapidly (see, for example, the first part of the experiment plotted in Figure 1—with animals that had no pretraining). A consequence of this is that a succession of three or four negative trials, as is bound to occur quite often in a random sequence, can result in the animals ceasing to take the test objects altogether for a period of hours or even days. During this time, octopuses subjected to a more evenly distributed sequence of positive and negative trials continue to learn; but the results of the two are scarcely comparable as the former animals presumably learn nothing about discrimination while refusing to take the objects. Unless a systematic order is used, the length of training is normally insufficient for such effects of trial sequence to cancel out and to ensure that all individuals are subject to comparable proportions of positive and negative trials. The sequences given above avoid strict alternation while ensuring that each animal has had the same number of positive and negative trials by the end of each day's training. In the 20-trial groups, the sequence is further arranged so that each half of each group contains 5+ and 5− trials; learning within the group and the effect of intensive training on the overall level of response can be studied (see Wells, 1962). For the present experiments, 18 octopuses were trained, but one of these was later found to have some optic nerves remaining, and the results with this animal have been discarded.

The objects to be distinguished were of perspex, P1 and P4 being respectively grooved and smooth cylinders, PC a cube, PC2 a similar cube with the corners rounded off, and PS a sphere. A sixth object, PR, was a length of perspex rod having the same radius of curvature as the rounded corners on PC2. The objects are shown, drawn to the same scale, in Figures 1 and 2; the cylinders were 3·0 cm. long and 2·5 cm. in diameter. No effort was made to ensure that they were of the same weight as previous experiments had shown that octopuses cannot detect weight differences (Wells, 1961*a*).

## RESULTS

Nine out of the seventeen animals were first trained, at eight trials per day, to distinguish between P1 and P4, with the smooth cylinder P4 as positive object. This is an easy discrimination for octopuses (Wells and Wells, 1957) and they all learned to discriminate. All 17 animals were then trained to distinguish between PS and PC, eight of them with PS as positive object and nine with PC as positive. These results are summarized in Figure 1; most of the animals failed to discriminate

FIGURE 1



A summary of the performance of octopuses trained to discriminate by touch between a rough and a smooth cylinder (P1 and P4, shown above) and then between a sphere and a cube (PS and PC). ● shows the number of times the positive and ○ the number of times the negative object was taken. There were 8 trials per day, at intervals of not less than 1 hr.

at all, and in no case was performance as good as that of the worst animal trained on the textural (P1/P4) discrimination. Errors were mainly by acceptance of the negative object; there was no detectable difference in performance of the animals with and without previous training on P1/P4, and the data from the two groups have been added together in compiling Figure 1.

The rate of training was then raised from 8 to 40 trials per day, and the experiment continued under these conditions with 14 of the animals (7 with PC$^+$, 7 with PS$^+$). The object of increasing the rate of training was to reduce the tendency to take all of the objects presented, previous experiments having shown that a capacity to discriminate is sometimes masked by a preponderance of positive responses (Wells and Wells, 1958). Increasing the rate of training (and thus the frequency of shocks for erroneous positive responses) decreased the proportion of objects taken as expected, and in the last 20 of 80 trials at the new rate 8 of the 14 animals made 75 per cent. or more correct responses. Details of individual performances at this stage in the experiment are given in Table I.

TABLE I

RESULTS OF TRAINING TO DISCRIMINATE BETWEEN A CUBE AND A SPHERE, AND TRANSFER EXPERIMENTS

| Animal | | H1 | | | H2 | | | H3 | | | H4 | | | H5 | | | H9 | | | H10 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PS$^+$ | | | PS$^+$ | | | PS$^+$ | | | PS$^+$ | | | PC$^+$ | | | PC$^+$ | | | PC$^+$ | | |
| | | + | − | e | + | − | e | + | − | e | + | − | e | + | − | e | + | − | e | + | − | e |
| PS/PC Last 4 days of training | | 10<br>10<br>10<br>9 | 10<br>10<br>10<br>6 | 10<br>10<br>10<br>7 | 10<br>7<br>10<br>10 | 7<br>4<br>4<br>5 | 7<br>7<br>4<br>5 | 6<br>8<br>7<br>8 | 1<br>5<br>1<br>2 | 5<br>7<br>4<br>4 | 10<br>10<br>10<br>9 | 7<br>7<br>8<br>0 | 7<br>7<br>8<br>1 | 10<br>10<br>10<br>10 | 10<br>8<br>9<br>2 | 10<br>8<br>9<br>2 | 10<br>10<br>10<br>9 | 10<br>10<br>6<br>4 | 10<br>10<br>6<br>5 | 10<br>10<br>10<br>10 | 10<br>10<br>6<br>7 | 10<br>10<br>6<br>7 |
| Training continues | 1. With $PC_2$ substituted for PC | — | | | 8<br>8 | 8<br>4 | 10<br>6 | 8<br>5 | 1<br>1 | 3<br>6 | 6<br>3 | 3<br>1 | 7<br>8 | 9<br>10 | 4<br>4 | 5<br>4 | 6<br>9 | 1<br>5 | 5<br>6 | 10<br>10 | 8<br>8 | 8<br>8 |
| | 2. With PR substituted for PS or $PC_2$ | — | | | $PR^+/PC_2^-$<br>5<br>$PS^+/PR^-$<br>6 | <br>4<br><br>6 | <br>9<br><br>10 | $PS^+/PR^-$<br>10<br>$PR^+/PC_2^-$<br>4 | <br>7<br><br>8 | <br>7<br><br>14 | $PS^+/PR^-$<br>10<br>$PR^+/PC_2^-$<br>6 | <br>4<br><br>8 | <br>4<br><br>12 | $PR^+/PS^-$<br>10<br>$PC_2^+/PR^-$<br>0 | <br>5<br><br>2 | <br>5<br><br>12 | — | | | — | | |
| Retraining on PS/PC | | — | | | — | | | 10<br>10 | 7<br>3 | 7<br>3 | 10<br>10 | 8<br>5 | 8<br>5 | — | | | — | | | — | | |
| Training continues after damage to the inferior frontal system | | — | | | — | | | Operated<br>5<br>4 | <br>5<br>5 | <br>10<br>11 | <br>6<br>5 | <br>4<br>5 | <br>8<br>10 | — | | | — | | | — | | |

| Animal | | H11 | | | H12 | | | H16 | | | H21 | | | H23 | | | H25 | | | H28 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PC$^+$ | | | PC$^+$ | | | PC$^+$ | | | PS$^+$ | | | PS$^+$ | | | PC$^+$ | | | PS$^+$ | | |
| | | + | − | e | + | − | e | + | − | e | + | − | e | + | − | e | + | − | e | + | − | e |
| PS/PC Last 4 days of training | | 10<br>10<br>10<br>10 | 10<br>8<br>6<br>8 | 10<br>8<br>6<br>8 | 3<br>6<br>5<br>9 | 4<br>3<br>4<br>5 | 11<br>7<br>9<br>6 | 10<br>10<br>10<br>10 | 10<br>10<br>9<br>10 | 10<br>10<br>9<br>10 | 10<br>10<br>10<br>10 | 2<br>1<br>0<br>0 | 2<br>1<br>0<br>0 | 10<br>10<br>10<br>8 | 6<br>5<br>3<br>1 | 6<br>5<br>3<br>3 | 10<br>10<br>10<br>10 | 10<br>10<br>10<br>9 | 10<br>10<br>10<br>9 | 10<br>10<br>10<br>10 | 10<br>8<br>7<br>3 | 10<br>8<br>7<br>3 |
| Training continues | 1. With $PC_2$ substituted for PC | — | | | 5<br>3 | 5<br>4 | 10<br>11 | — | | | 4<br>7 | 0<br>0 | 6<br>3 | 5<br>3 | 3<br>1 | 8<br>8 | — | | | 10<br>10 | 10<br>6 | 10<br>6 |
| | 2. With PR substituted for PS or $PC_2$ | — | | | — | | | — | | | $PR^+/PC_2^-$<br>0<br>$PS^+/PR^-$<br>6 | <br>0<br><br>0 | <br>10<br><br>4 | $PR^+/PC_2^-$<br>0<br>$PS^+/PR^-$<br>10 | <br>1<br><br>3 | <br>11<br><br>3 | — | | | $PR^+/PC_2^-$<br>1<br>$PS^+/PR^-$<br>10 | <br>1<br><br>0 | <br>10<br><br>0 |
| Retraining on PS/PC | | — | | | — | | | — | | | 5<br>10 | 1<br>1 | 6<br>1 | 10<br>10 | 5<br>3 | 5<br>3 | — | | | 10<br>10 | 8<br>2 | 8<br>2 |
| Training continues after damage to the inferior frontal system | | — | | | — | | | — | | | Operated<br>4<br>8 | <br>5<br>8 | <br>11<br>10 | <br>6<br>9 | <br>5<br>8 | <br>9<br>9 | — | | | Operated<br>7<br>7 | <br>7<br>8 | <br>10<br>1 |

The columns + − and e show, respectively, the number of times that each object was taken and the total errors in each group of 10 trials.

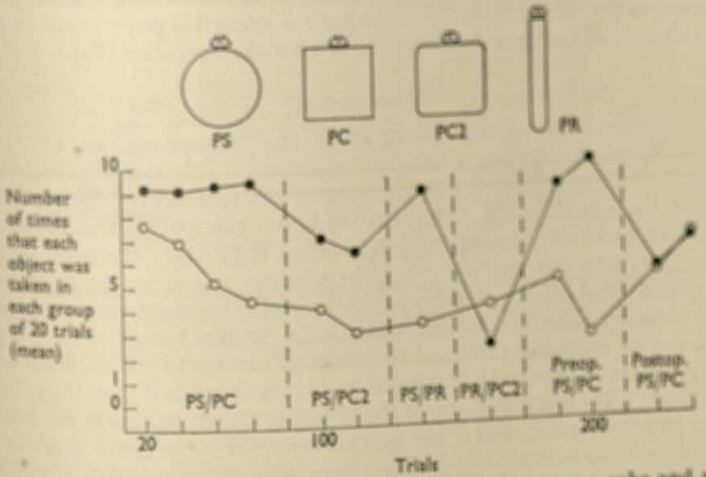
It will be noticed that the octopuses were now making a tactile shape discrimination, which previous experience had indicated should be impossible for them. The rest of the experiment was devoted to discovering what attribute or attributes of cube or square they were using to distinguish between them.

In considering possible differences between PS and PC, it must be remembered that an octopus grasps things that it handles with suckers, each of which is cup shaped, with a circular rim. Applied to a smooth surface, the rim is in contact but the rest of the sucker may not be touching the object at all; sense organs are concentrated in the rims of the suckers (Rossi and Graziadei, 1958). So far as any individual sucker is concerned, therefore, a section of a sphere must be indistinguishable from a plane surface. In order to distinguish between a sphere and a cube, each several times the diameter of an individual sucker, the octopus must therefore *either* be able to locate the relative positions of the individual suckers in contact with the object (which implies specification of how the arm to which the suckers are attached is bent), *or* be able to recognize that individual suckers are distorted.

The experiments that follow show that *Octopus* distinguishes between cubes and spheres by recognizing that the cube has corners that bend suckers, while the sphere has not.

It is not, of course, possible to make a cube without corners, but the corners can be made less well defined. This was done in the case of the cube PC2, which was like PC, but had the sharp corners rounded off (Fig. 2). When PC2 was substituted for PC, to animals already trained to distinguish between PC and PS continued to discriminate, but made a higher proportion or errors than before the change; there were only 63 per cent. correct responses made in the 40 trials following the introduction of PC2, compared with 79 per cent. correct responses in the 40 trials immediately preceding it ($\chi^2$ 1 d.f. = 18·3, $p < 0·01$) (Fig. 2 and Table I).

FIGURE 2



Summarizes the last 80 trials of training to distinguish between a cube and a sphere, and the effect of substituting further objects for one or other of these. For details of the order in which the tests were carried out in individual animals, see Table I. As before, ● shows the number of times the positive object, and ○ the number of times the negative object was taken. There were 40 trials per day in two groups of 20 (each 10+, 10−). Finally the animals were retrained and lesions made to the inferior frontal system, after which discrimination broke down.

Training was then continued with seven of the animals (the other three were used for an experiment on size discrimination), with a further object, PR, substituted for PC2 or PS. PR was a length of perspex rod having the same radius of curvature as the rounded corners of PC2. In two instances PR was first presented in place of PC2 and in five, in place of PS. Twenty training trials were made, followed by 20 more trials with PR substituted for the other object. The results of these experiments are given in Figure 2 and Table I; they may be summarized as showing that PR is treated as a cube. Thus in the six instances where PR replaced the sphere as positive object, errors rose sharply because the animals, already trained to reject PC2, rejected PR as well; these six octopuses averaged 11·7 errors in 20 trials, compared with 6·3 out of 20 before the change ($\chi_2$ 1 d.f. = 15·6 $p < 0·01$). When PR was presented instead of PC2, performance actually improved, and the animals made rather *fewer* errors than before the substitution, averaging only 4·7 mistakes (cf. 6·3, the difference is not, however, significant —$\chi_2$ 1 d.f. = 2·03, $p$ 0·15). A similar result was obtained in the single instance where PC2 was positive object before the change. This animal made four errors in the 20 trials before PR was substituted for PC2. In the next 20 trials, with PR in place of PC2, it did still better, making only three mistakes. When training was continued, with PR given in place of PS, the negative object, the animal started making mistakes by taking PR as well as PC2; despite frequent shocks it made 12 mistakes in the following 20 trials. Clearly PR was not only confused with PC2 in these experiments; it was, if anything, a "better" cube than PC2 so far as the octopuses were concerned, presumably because it had no flat surfaces that could be mistaken for sections of the sphere.

Finally five of the animals had further training to distinguish between PS and PC before having the inferior frontal system damaged in brain operations intended to destroy parts already known to be concerned in learning to distinguish between textures by touch. The forms of these lesions are shown in Figure 3 and the performance of the octopuses before and after the operation summarized in Figure 2; the animals did not continue to discriminate after the operation (Table I).

## Discussion

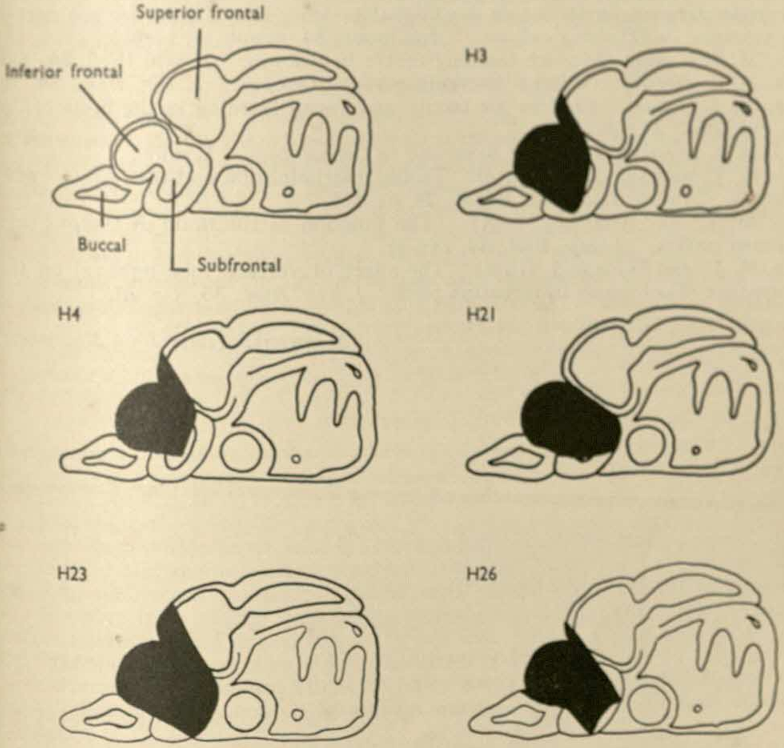
The experiments described above show that octopuses can be taught to distinguish by touch between a cube and a sphere. This was unexpected. They also show that the basis of this discrimination is the presence or absence of corners; if these are made less abrupt discrimination is less efficient; if the flat surfaces of this modified cube are eliminated, performance improves. Rounded corners are evidently more like flat surfaces than sharp ones, and it would in principle be possible from a series of such experiments to determine the maximum radius of curvature that *Octopus* is able to distinguish from a flat surface.*

A rod having the same radius as the rounded corners of the modified cube was not only mistaken for the cube, but more readily distinguished from the sphere than the cube itself. The overall shape of the object is therefore clearly not important, compared with the presence or absence of corners. It is the latter that constitutes the detectable difference between a cube and sphere so far as *Octopus* is concerned. It is not therefore necessary to suppose that the animal is using information about the relative positions of the individual suckers, or about the way that the arm is bent in wrapping around the object; it need only be supposed that it can detect that the rims of some of the suckers are distorted by the contact. The present experiments thus confirm previous work, all of which has gone to show that octopuses are incapable of utilizing information from internal receptors in learning (for a survey and discussion of these experiments, see Wells, 1962).

---

* Since this was written (April, 1962) experiments with rods of differing diameter have shown a clear relation between difference in surface curvature and the number of errors in learning to discriminate; moreover units made by sticking narrow rods together in bundles are treated as being of the diameter of their components, and not in relation to their overall diameter. These results are consistent with the theory that what matters is the distortion imposed on individual suckers, rather than the curvature imposed on the arm as a whole (Wells, *J. exp. Biol.* (in press)).

FIGURE 3



Maps of lesions made in animals trained to distinguish between a cube and a sphere. The part removed in each case is shown in black on a diagram representing a median longitudinal section through the supraoesophageal lobes of the brain. None of these animals continued to discriminate after the operation.

It is not known what receptors are concerned in registering that the rim of a sucker is bent, but a number of types of sensory cell have been described from the epithelium and superficial muscular layers of the rims of the suckers (Rossi and Graziadei, 1958) and several of these are so placed that they must be subjected to compression, or shear forces, when the rim is bent. Discrimination seems likely to be based on signals from these receptors, which are in all probability also the receptors concerned in the discrimination of objects differing in texture; experiments on textural discrimination indicate that the amount of deformation of the suckers in contact is the quantity measured by the animal (Wells and Wells, 1957).

The breakdown of discrimination following inferior frontal damage shows at least that whatever the source of the information on which such discriminations are based, the same filtering and storage mechanisms are concerned in learning cube/square discriminations as in learning to distinguish between objects of differing texture (Wells, 1959, 1961*b*, 1962).

The author would like to thank the Director and staff of the Naples Zoological Station for the facilities that they provided, and Professor J. Z. Young, F.R.S., who kindly read and commented on the manuscript in preparation.

## REFERENCES

ROSSI, F., and GRAZIADEI, P. (1958). Nouvelles contributions à la connaissance du système nerveux du tentacule des Céphalopodes. IV. Le patrimonie nerveux de la ventouse de l'*Octopus vulgaris*. *Acta. anat.*, **34**, suppl. 32, 1–79.

WELLS, M. J. (1959). A touch-learning centre in *Octopus*. *J. exp. Biol.*, **36**, 590–612.

WELLS, M. J. (1961*a*). Weight discrimination by *Octopus*. *J. exp. Biol.*, **38**, 127–33.

WELLS, M. J. (1961*b*). Centres for tactile and visual learning in the brain of *Octopus*. *J. exp. Biol.*, **38**, 811–26.

WELLS, M. J. (1962). *Brain and Behaviour in Cephalopods*. London.

WELLS, M. J., and WELLS, J. (1956). Tactile discrimination and the behaviour of blind *Octopus*. *Pubbl. Staz. zool. Napoli*, **28**, 94–126.

WELLS, M. J., and WELLS, J. (1957). The function of the brain of *Octopus* in tactile discrimination. *J. exp. Biol.*, **34**, 131–42.

WELLS, M. J., and WELLS, J. (1958). The effect of vertical lobe removal on the performance of octopuses in retention tests. *J. exp. Biol.*, **35**, 337–48.

*Manuscript received 5th November, 1962.*

# SHORTER ARTICLES AND NOTES

## PROBABILITY-LEARNING IN THE MONKEY

BY

WM. A. WILSON, JR., MARLENE OSCAR, and M. E. BITTERMAN

*From the Department of Psychology, Bryn Mawr College*

Tested in a choice-situation, two rhesus monkeys showed a correspondence between choice-ratios and reinforcement-ratios. The basis of this event-matching proved to be different in the monkey from that in lower forms.

### INTRODUCTION

Previous experiments on probability-learning in the monkey (Wilson and Rollin, 1959; Wilson, 1960; Meyer, 1960) have given no evidence of the event-matching which is found in fish (Behrend and Bitterman, 1961) and pigeon (Bullock and Bitterman, 1962). Trained in choice-situations, both fish and pigeon have been found to show a close correspondence between choice-ratio and reinforcement-ratio as the reinforcement-ratio is varied from 50 : 50 to 100 : 0. If, for example, one alternative is reinforced on a random 70 per cent. of trials and the other on the remaining trials, both species come to choose the 70 per cent. alternative on about 70 per cent. of trials. By contrast, several experiments with the monkey have yielded only maximizing—that is, choice of the more frequently reinforced alternative (whatever the ratio of reinforcement) on almost all trials. In the experiment to be reported here, a correspondence between choice-ratio and reinforcement-ratio was found for the first time in monkey.

### METHOD

Two experimentally naïve rhesus monkeys, 2–3 years old, were tested in a chamber which was set into a ventilated, sound-reducing enclosure. In one wall of the chamber were mounted two retractable levers (7 in. apart and 12 in. above the floor); a centred food-cup (5 in. above the floor); and a centred Plexiglas panel (2 in. × 3 in.) at the height of the levers. Each session began with the levers retracted and the chamber dark. After a 10-sec. interval, the panel was illuminated from the rear; 10 presses on the panel turned the panel-light off, turned the house-light on, and caused the levers to be presented. Response to either lever caused both levers to retract and turned off the house-light. An incorrect response was followed simply by a 10-sec. period of darkness, after which the panel-light was turned on once more. A correct response delivered a peanut to the food-cup and turned on the food-cup light for 4 sec.; a 6-sec. period of darkness intervened between the offset of the food-cup light and the onset of the panel-light with which the next trial began.

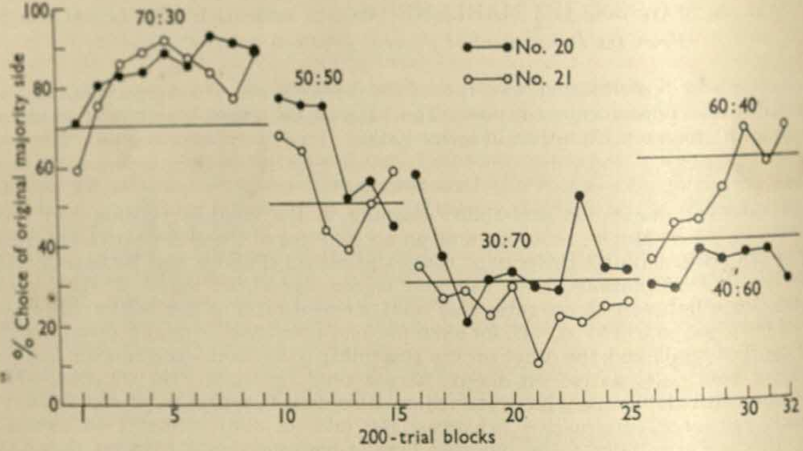
The animals were tested in 25-trial sessions, of which two usually were given each day. The reinforcement-ratio was 70 : 30 for 1,800 trials, then 50 : 50 for 1,200 trials, and 30 : 70 for 2,000 trials. For the last 1,400 trials, the ratio was 40 : 60 for one animal (No. 20) and 60 : 40 for the other (No. 21). Random series were chosen to determine which lever would be correct on a given trial, with the restriction that each lever be correct at least once in each block of five trials and that the predetermined reinforcement-ratio be met exactly in each set of 50 trials. A correction procedure was used: no limit was placed on the number of repetitive errors which could be made, and a trial was defined as a series of choices ending with a correct response. All of the experimental contingencies were programmed automatically. The responses of the animals were recorded on tape.

### RESULTS

Plotted in Figure 1 are the individual preferences in each phase of the experiment for the lever more frequently reinforced in the first phase. The curves are based only on the first choice made on each trial, repetitive errors and correct choices which followed errors being omitted from the computations. (Consideration of first-choices alone is dictated by the fact that the correction method forces response to the minority stimulus.

When all responses are counted, the results are biased against maximizing; pure maximizing cannot, in fact, appear. The bias becomes especially strong with the elimination of repetitive errors as training continues.) As may be seen in the Figure, both animals tended to maximize in the first phase of the experiment, but in the subsequent phases their choice-ratios came closely to approximate the reinforcement-ratios.

FIGURE 1



Individual preferences in each phase of the experiment for the lever more frequently reinforced in the first phase. The reinforcement-ratio in each phase is shown.

Several questions may be asked with respect to these data:

(1) What is the reason for the discrepancy between the results for the first phase of the experiment and those for the subsequent phases? Experience with the situation may be important; it is conceivable, for example, that further training on the initial 70:30 problem would have brought the curves down to the 70 per cent. level. It is conceivable, too, that the 50:50 schedule served to break up a strong initial bias. The first question can be answered only by further experiments.

(2) What is the reason for the discrepancy between the results of the last three phases of this experiment and the results of earlier experiments with the monkey? In some of the previous work, a noncorrection procedure was used; this procedure, which does not ensure that the animal will be reinforced for choice of the lower-probability alternative, also produces maximizing in fish and pigeon. In other previous work with the monkey, reliance was placed on "implicit" correction, which, like noncorrection, fails to control the distribution of reinforcement. Only in one previous experiment, which involved guidance for some monkeys and a form of correction for others, did the procedures employed resemble those which produce matching in fish and pigeon, but training was not carried beyond an initial 75:25 problem. Whether in that experiment, as in the present one, further training on other problems would have produced matching cannot be said. It should be noted, however, that the conditions of the present experiment were selected as more likely than those of the earlier one to show matching (if it was to be found at all in the monkey) on the basis of a study of the conditions which make for matching in the pigeon (Graf, Bullock, and Bitterman, 1964). One rule which emerges from this study is that matching is more likely to appear when the consequences of incorrect choice are of substantial motivational significance.

(3) What is the relation between the matching which appears in the last three phases of this experiment and the matching of fish and pigeon? That the two

> phenomena are quite different is suggested by the fact that both monkeys showed a marked tendency toward alternation of response after reinforcement. In the 50:50 problem of the second phase, alternation of choice after reinforcement was more than three times as likely as persistence. In the 30:70 problem of the third phase, the alternating tendency was superimposed on a clear preference for the majority stimulus; the animals chose the 70 per cent. lever on about 96 per cent. of trials following reinforcement on the 30 per cent. side, as compared with 58 per cent. of trials following reinforcement on the 70 per cent. side $(0{\cdot}96 \times 0{\cdot}30 + 0{\cdot}58 \times 0{\cdot}70 = 0{\cdot}69)$. The 40:60 and 60:40 problems of the fourth phase produced like results. No such systematic tendency can be found in the data for fish and pigeon; at asymptote in a 70:30 problem, for example, the 70 per cent. alternative is chosen on about 70 per cent. of trials irrespective of which alternative has been reinforced on the trial immediately preceding.

Experiments with the rat under conditions analogous to those which produce matching in fish and pigeon have for the most part yielded only maximizing. One recent experiment (Hickson, 1961) did, however, produce a correspondence between choice-ratio and reinforcement-ratio which could be traced to a strong tendency of the animals to choose on each trial the alternative reinforced on the preceding trial—an outcome better classed with the results for monkey than with those for fish and pigeon. (This statement is based on Bitterman's analysis of a sample of Hickson's data which was provided by Hickson). It is possible, of course that some change of procedure will produce in rat or monkey the kind of unsystematic matching which can be demonstrated in fish and pigeon, but evidence now available seems to indicate the operation of different mechanisms of choice in mammalian and submammalian forms.

This research was supported in part by Grant M-4472 from the U.S. Public Health Service and in part by Contract Nonr 2829(01) with the Office of Naval Research.

## References

Behrend, E. R., and Bitterman, M. E. (1961). Probability-matching in the fish. *Amer. J. Psychol.*, **74**, 542–51.

Bullock, D. H., and Bitterman, M. E. (1962). Probability-matching in the pigeon. *Amer. J. Psychol.*, **75**, 634–9.

Graf, V., Bullock, D. H., and Bitterman, M. E. (1964). Further experiments on probability-matching in the pigeon. *J. exp. Anal Behav.*, **7**, 151–7.

Hickson, R. H. (1961). Response probability in a two-choice learning situation with varying probability of reinforcement. *J. exp. Psychol.*, **62**, 138–44.

Meyer, D. R. (1960). The effects of differential probabilities of reinforcement on discrimination learning by monkeys. *J. comp. physiol. Psychol.*, **53**, 173–5.

Wilson, W. A., Jr. (1960). Two-choice behavior of monkeys. *J. exp. Psychol.*, **59**, 207–8.

Wilson, W. A., Jr., and Rollin, A. R. (1959). Two-choice behavior of rhesus monkeys in a noncontingent situation. *J. exp. Psychol.*, **58**, 174–80.

*Manuscript received 15th September, 1962.*

# MODIFICATION OF THE DIVING-IN RESPONSE OF THE RED-EARED TERRAPIN, *PSEUDEMYS ORNATA CALLIROSTRIS*

BY

N. MROSOVSKY

*From the Medical Research Council Unit for the Experimental Investigation of Behaviour, University College London*

A terrapin will dive off a platform into water. This response was used in a simple apparatus with alternative water compartments to study discriminations. A white side was chosen in preference to a black side. By introducing shock it was possible to reverse this or, in some animals, inhibit diving-in. The variability of terrapins and the limitations and advantages of the present method for training them are considered.

## INTRODUCTION

It is not easy to devise satisfactory training methods for chelonians. Their wide range in growth rate and their ability to go without food for long periods make hunger motivation difficult to use. Their reaction to shock is to withdraw into their shells so they are not obviously suitable for escape training. However there have been a number of studies with food incentive, or food and shock combined, which have shown that chelonians can discriminate colours and improve in mazes, even if in some cases it has been necessary to "discard any turtle that refuses food, retires within its shell upon the slightest provocation and shows a pronounced disposition to sulk" (Casteel, 1911). The literature on training experiments is summarized by Boycott and Guillery (1962). Nearly all investigators mention great individual differences.

The uncertainties inherent in using food motivation have been stressed by Yerkes (1901). He found that a comfortable nest was sufficient reward to reduce the time and error scores of a speckled tortoise, *Chelopus guttatus*, in a simple maze. The present paper describes an attempt to produce learning and consistent performance and some of the problems met using entry into water as motivation. Recently Gonzalez and Bitterman (1962) used water reward with *Chrysemys picta picta* and scored on the basis of time. The present experiment was designed to use water entry as an incentive to form a discrimination and score according to the choice made.

## METHOD

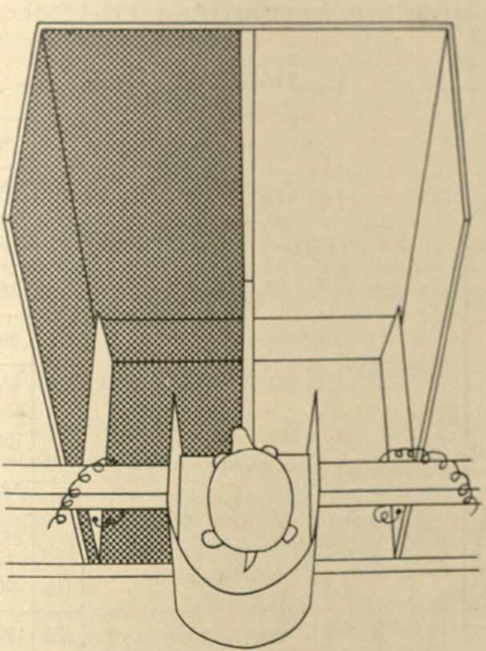
*Subjects and conditions of housing.* The subjects were 19 red-eared terrapins, *Pseudemys ornata callirostris* (Gray). On arrival from North America in June their weights ranged from 4 to 8 gm., and the umbilicus was not completely sealed which suggests they may only have been out of the nest a few weeks. Their age when the experiment was begun was probably about 5 months. The conditions of housing and the diets were similar to those described by Boycott and Robins (1961), except that very little ultra violet light was given. The temperature in the home tanks was held at 26 ± 6° C. Weights of the animals were recorded weekly.

*Experimental procedure.* Use was made of the fact that these terrapins when placed on a platform or rock scuttled off into water below. They would do this either when put on a small stand, as was done in the training situation, or if sitting on the rock in their home tank when someone approached or sudden stimuli occurred. Although kept in the laboratory many months there were no signs of increasing tameness or habituation to stimuli that produced entry into water.

*Apparatus.* The apparatus is shown in Figure 1. Access to water from a grey stand of about 200 sq. cm. was possible by diving down into either a matt black or matt white compartment. The base of each compartment was 12 × 14 cm. and, at the two sides, walls extended up to a height of 36 cm. The depth of the water was 7 cm., and its surface 5 cm. below the platform. The temperature was held at 26 ± 3° C. to avoid giving the animals thermal shock.

*Shock levels.* With electrode plates in the water it is difficult to equate the shock strengths for different animals. An A.C. transformer supplied current across the plates. But if an animal is put into the water the current that flows through it will depend on its size and orientation. If its extremities are far apart the potential difference across them will be relatively high. Terrapins reacted more strongly when facing the plates, than if lying lengthwise between them, and this was presumably because the potential difference across them was greater when their longest axis was at right angles to the electrodes. To measure and equate currents flowing across each animal was not attempted, but the levels across the electrodes varied after consideration of the performance of individuals at a given strength. The current strengths stated refer to what passed across the electrodes. Since the water takes most of this the current actually flowing across the animal will be far less, in the order of a few milliamps.

FIGURE 1



Terrapin training apparatus. The two other electrode plates at the division between the water compartments are not shown in the figure.

With very weak 2 sec. duration shocks a terrapin may show increased activity. At levels of 260 ± 20 milliamp. across the plates, which were initially used in this experiment, most animals showed head retraction and immobility during the shock. Later levels of 510 ± 20 millamp. were introduced for animals not showing further progression in learning. For one very small terrapin, 2R, 710 ± 20 milliamp. was used. At these higher strengths the limbs may be extended and sometimes show transient tremors after the shock has ended; a period of reduced activity and diminished reaction to startle stimuli may follow.

*Training schedule.* Four trials a day were given on 5 days of the week. There was no training at the week-end. On each day the black shaft was twice on the right and twice on the left, the order being random. For the first 3 weeks (60 trials) no shocks were given. After this the animals were divided into two groups. For 5 animals the no shock procedure was continued. For the other 14 animals 2-sec. duration shocks were given as they touched the water on the white side. The terrapin was picked out of the home tank by the experimenter, who wore a light brown coat, and placed on the platform. The experimenter stood behind the platform and recorded the time to enter the water.

If this exceeded 5 min. the trial was ended. If the animal entered the water 5 sec. elapsed before it was removed and returned to its home tank.

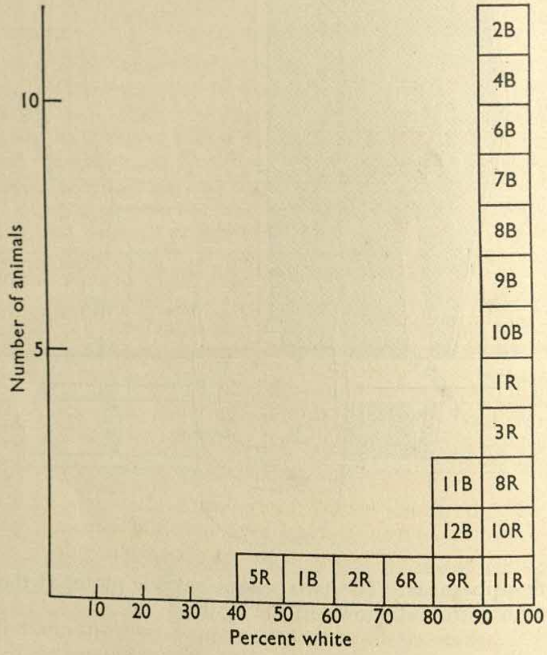
The animals were fed in the morning, and the training took place between 2.0 and 5 p.m. The intertrial interval was not constant but was usually 10 or 20 min.

Retention tests were run on eight subjects at the end of the experiment. Shocks were used in retesting.

## RESULTS

*Preferences.* Figure 2 shows the strong white preference of most animals. On the very first trial 4 terrapins that later stabilized on the white choice went to the black side. The white preference became even more pronounced in most animals as training progressed. In 2 animals, 5R and 1B, there was no definite preference, nor did this appear when non-shock trials were extended up to totals of 140 and 100 respectively. Subsequent training and their reaction in an optomotor apparatus suggested that their eyesight was adequate.

FIGURE 2



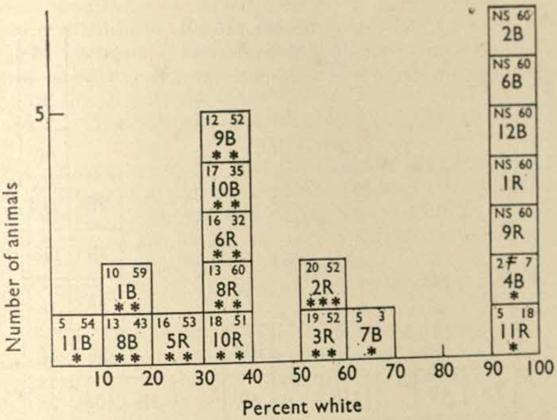
Percentage to the White of Responses made in the 3 weeks of preference testing.

Terrapins sometimes showed slight right and left preferences. Some animals would turn round on the platform at times rather than go straight to the water. But position habits were heavily overshadowed by the white preference in all animals except possibly 2R and 5R.

During the first week or two of preference testing the responses were less rapid (some animals tending to freeze) and differed more between animals. But by the third week most animals were diving within a few seconds. This low latency (usually $< 5$ sec.) remained consistent in four of the animals kept without shock for 20 ± 1 weeks. But the behaviour of one terrapin (2B) was different in that during the first 3 weeks of training it only dived on 57 per cent. of the trials in the 5 min. allotted whereas all the other animals went in more than 95 per cent. of the time. But with continued trials its behaviour became indistinguishable from that of other unshocked animals (see Fig. 4).

*Learning.* Comparison of Figures 2 and 3 shows the decrease of responses to the white side with the introduction of suitable shock levels. Figure 4 shows that in some cases responses to the white are replaced by remaining on the platform, e.g. 7B, 4B and 11R. These 3 animals moved around on the platform and retreated from its edges. The 20 successive trials they made without entering the water cannot be attributed to the reduced activity that sometimes follows shock. In other animals, e.g. 6R and 10B, it is hard to be sure whether the staying still on the platform was simply a direct effect of shock or learning that entry into the water was liable to be punished. But learning is demonstrable from their ratio of black to white responses (compare Figs. 2 and 3).

FIGURE 3



Percentage to the White of Responses made for the last 3 weeks of training
Maximum shock levels:

* = 260 milliamp.
** = 510 milliamp.
*** = 710 milliamp.
NS = No shocks given.
‡ = based on 2 weeks training only.

In the centre of each square the number of the Terrapin is given. Figures in the top left show the total number of weeks training (60 trials per week) after the preference testing ended; top right shows the number of responses on which the percentage White for the last 3 weeks is based.

*Retention.* Out of 8 terrapins tested for retention after gaps of 4–6 weeks without training, one showed no change from the last week of initial learning; seven showed slight decrements but still went more to the black than to the white side.

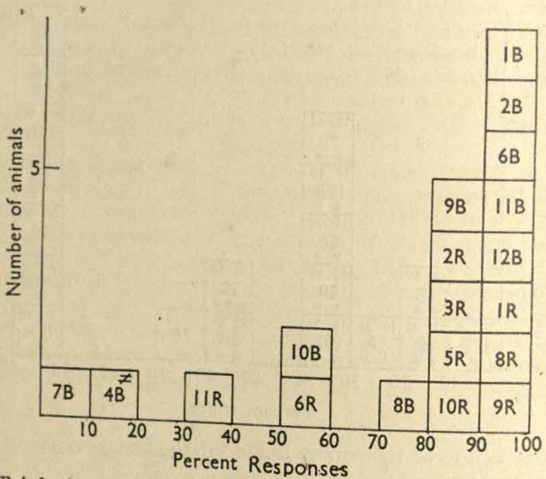
## DISCUSSION

It seems unlikely that food motivation was operative although the animals were always fed by someone wearing a white coat. It is however possible that the preference was acquired while the animals were in the laboratory, which had off-white walls. But perhaps it is more probable that the white is an indication that water is likely to be present. For various turtles from the Pearl River area Anderson (1958) suggested that an open horizon was one of the factors that guided the hatchlings to water. The black shaft in the present apparatus may look too solid to be inviting to the terrapin. But the same preference may not necessarily occur in other species or under different motivations. Casteel (1911) found that in a Yerkes box with food and shock incentive *Chrysemys marginata* tended to show a black preference and suggested this was due to transfer from previous training. However Noble and Breslau (1938) found that several species of fresh-water turtles shortly after hatching were attracted to areas of intense illumination.

The results of this experiment show that the diving-in response of *Pseudemys ornata callirostris* can be used as the basis of a training method. The procedure has similarities to the Lashley jumping stand for rats in that the stimulus should be in the visual field as the response is made. One advantage of this apparatus is that with most terrapins lengthy pretraining is unnecessary. The performance of aquatic turtles on a visual cliff where only 76 per cent. choose the shallow side, compared to 95–100 per cent. with rats, is perhaps a related phenomenon. This could be due to lack of fear rather than poor depth perception (Gibson and Walk, 1960).

There are several problems relevant to training the terrapin that the present experiment raises. First if the preferences become stronger as more trials are given it is doubtful that the word preference is entirely appropriate. Kirk and Bitterman (1963) have also found that in *Chrysemys picta picta* the preference ultimately displayed, a black-right one, was not the same as that shown on the first trial.

FIGURE 4



Percentage of Trials for the last 3 weeks of Training on which a response was made within 5 min.

‡ = based on only 2 weeks training.

The presence of the experimenter behind the stand and the generally unquiet conditions of the laboratory certainly influenced the latency. There were great individual differences in latency both before and after the introduction of shock. Casteel (1911) with his maze found time scores to be of no use. Gonzalez and Bitterman (1962) found them satisfactory enough to make them the basis of their method. In the present experiments the time score is of limited value.

The difficulty in finding the best shock level is the main limitation in the present method. With weaker shocks there is little indication of learning in some animals but at higher settings response times are so long that trials have to be abandoned. Casteel (1911) also noted inertness in some animals following shock. Another point that must be considered in using shock and food motivation in combination is that in many of the terrapins there was a reduced growth rate following the introduction of shock. Further data on this will be presented elsewhere (Edwards and Mrosovsky, 1964).

As regards the variability in learning it must be remembered that the animal can avoid the white either by remaining on the platform or by going to the black. Which method is adopted appears to some extent to be related to the strengths of the initial preferences (e.g. 4B and 1B). It is surprising though that 5R and 2R only showed slight and unstable signs of learning, for if the task involves the suppression of a preferred response, one would have expected these animals to have scored better since their initial preferences were weak.

However the differences in shock used and in reaction to shock make it difficult to compare learning between animals. But it is clear that chelonians can learn rapidly either to make a discrimination (11B) or not to respond (11R, 7B and 4B). It seems that all animals have shown learning of one of these types because the preference tests for the first 3 weeks provide a good base line for using the animal as its own control. The scores of animals kept on continuous preference testing indicate that the increase in black responses of animals receiving shock for going to the white is not due to a maturation factor. The consistent performance of these animals over many weeks is the most encouraging finding of this experiment.

I would like to thank Miss A. A. Edwards for invaluable help, Mr. B. B. Boycott for his advice and encouragement and the Medical Research Council for supporting the work. I am grateful to Miss J. A. Cochrane of the British Museum (Natural History) for identifying the Pseudemyds.

## REFERENCES

ANDERSON, P. K. (1958). The photic responses and water-approach behaviour of hatchling turtles. *Copeia*, 211–5.

BOYCOTT, B. B., and GUILLERY, R. W. (1962). Olfactory and visual learning in the red-eared terrapin, *Pseudemys scripta elegans* (Wied). *J. exp. Biol.*, **39**, 567–77.

BOYCOTT, B. B., and ROBINS, M. W. (1961). The care of young red-eared terrapins (*Pseudemys scripta elegans*) in the laboratory. *Brit. J. Herpetol.*, **2**, 206–10.

CASTEEL, D. B. (1911). The discriminative ability of the painted turtle. *J. Anim. Behav.*, **1**, 1–28.

EDWARDS, A. A., and MROSOVSKY, N. (1964). Electric shock and reduced weight gain in the terrapin, *Pseudemys ornata callirostris*. *Brit. J., Herpetol.*, in press.

GIBSON, E. J., and WALK, R. D. (1960). The "visual cliff." *Sci. Amer.*, **202**, 64–71.

GONZALEZ, R. C., and BITTERMAN, M. E. (1962). A further study of partial reinforcement in the turtle. *Quart. J. exp. Psychol.*, **14**, 109–12.

KIRK, K. L., and BITTERMAN, M. E. (1963). Habit reversal in the turtle. *Quart. J. exp. Psychol.*, **15**, 52–7.

NOBLE, G. K., and BRESLAU, A. M. (1938). The senses involved in the migration of young fresh-water turtles after hatching. *J. comp. Psychol.*, **25**, 175–93.

YERKES, R. M. (1901). The formation of habits in the turtle. *Pop. Sci. Mon.*, **58**, 519–25.

*Manuscript received 2nd October*, 1962.

# GROUPING OF STIMULI IN PATTERN RECOGNITION AS A FUNCTION OF AGE

BY

P. M. A. RABBITT

*From the Medical Research Council Applied Psychology Research Unit, Cambridge*

Two groups of subjects aged 20–40 and 40–60 years learnt identifying responses to a set of eight patterns. The groups did not differ in rate of learning, but an analysis of confusions between patterns revealed systematic differences in the extent to which rules underlying the construction of patterns were used in recognizing them as members of sub-sets before identifying responses were mastered.

## INTRODUCTION

It has been clear for some time that the simple association of a "stimulus" and a "response" is an inadequate description of the processes of adaptive selection which Bartlett (1932) has shown to be characteristic of human memory. A comparison between the rules of this selection and computer codes is a fashionable analogy (Oldfield, 1954) which has yielded interesting experiments (Ross, 1961).

Attneave (1955; 1957) and Fitts *et al.* (1956) have shown that subjects learning to identify patterns code the data they perceive so as to select cues which are discriminative for a set of patterns with which they are concerned, and to ignore cues which are misleading or redundant.

For any given ensemble of patterns there may be a single optimal set of cues having minimum redundancy, or a number of such sets representing equally efficient systems of classification. Sets of patterns may be constructed so as to offer hierarchical systems of classification, so that two cues may have to be considered in identifying a pattern, the first identifying a sub-set of which the pattern is a member, and the second identifying the pattern within this sub-set. A subject optimizing his coding system by learning such conditional cues may be said to be mastering the rules of classification of the set of patterns. Since relevant cues cannot be recognized, nor irrelevant cues ignored until the subject has some knowledge of the set of patterns as a whole, the progress of mastery of the rules of classification will be reflected in confusions made between patterns before they are identified completely.

A series of experiments by Rabbitt (1961) showed that older subjects are less efficient than young subjects when required to respond to categories of stimuli. Further, older subjects were found to be less able to ignore irrelevant information in perceptual tasks. The following experiment was undertaken to discover whether these results could be generalized for coding processes in pattern identification.

## MATERIAL

Eight patterns were used. Each consisted of two groups of four adjacent rectangles drawn in Indian ink on white pasteboard cards. Each rectangle measured 1½ in. × 4 in. and the two groups of rectangles were 2 in. apart. In each pattern two of these rectangles were inked in and the remaining six left unfilled. The patterns are shown symbolically below. X represents a filled rectangle and O an unfilled one.

*One handed patterns*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | O | O | O | O | X | O | O | X |
| 2 | O | O | O | O | O | X | O | X |
| 3 | X | O | X | O | O | O | O | O |
| 4 | O | X | O | X | O | O | O | O |

*Two handed patterns*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | O | O | X | O | X | O | O | O |
| 6 | O | O | O | X | O | X | O | O |
| 7 | X | O | O | O | O | X | O | O |
| 8 | O | O | X | O | O | O | X | O |

The eight patterns divide naturally into two sub-sets of four patterns each. In one sub-set both filled rectangles were in the same group of four (one-handed). In the other sub-set one of the filled rectangles was in each group of four (two-handed).

### *Procedure*

Subjects were tested in groups of six, all members of a group being of the same age-range (20–40 or 40–60 years). The experimenter presented the patterns by holding up the cards one at a time for 5 sec. Intervals between successive presentations were of the order of 20 sec. and were occupied by subjects recording their responses on numbered protocols.

At the beginning of an experimental session subjects passively observed while the experimenter showed the eight patterns twice in random order and read aloud the responses assigned to them. The eight responses to be used, which were always "Bob," "George," "Arthur," "John," "Dick," "Peter," "Harry," and "Jim," were then copied down by the subjects. This was done to eliminate errors due to the use of unassigned responses, which would give no information about confusions between patterns. The eight responses were assigned to the patterns at random, assignments being varied between successive groups of subjects to eliminate persistent response-confusions from the summed data. When subjects had listed the required responses the patterns were presented singly, in random order, until each had been shown 13 times. After each presentation subjects recorded their responses, and the experimenter then read aloud the correct response.

At no time during the experiment was any comment passed which might indicate a possible scheme of classification for the patterns.

### *Subjects*

Subjects were 58 postmen. Twenty-nine were aged 20–40 years (mean 27·4) and 29 were aged 40–60 years (mean 47·1).

## Results

1. *Rate of learning of correct identifications.* Error scores for each successive block of 26 trials were obtained for each group (see Table I) and compared by a Mann-Whitney U test. No significant differences were found. The groups can thus be equated for rate of learning, providing a criterion of comparability directly related to this task.

TABLE I

Error Scores for Successive Blocks of 26 Presentations

| | *1st Quarter* | *2nd Quarter* | *3rd Quarter* | *4th Quarter* |
|---|---|---|---|---|
| Youngs . . | 461 | 392 | 337 | 267 |
| Olds . . . . | 507 | 426 | 341 | 306 |

2. *Error distributions*

A. *Overall randomness of errors.* Errors made by each group of subjects were summed in a 7 × 8 error matrix (see Table II) and tested by $\chi^2$ to examine the hypothesis that they were randomly distributed. The 20–40 age group gave a $\chi^2$ of 259 and the 40–60 group a $\chi^2$ of 125, for 55 degrees of freedom showing that errors were not made at random by either group ($p < 0{\cdot}001$).

This procedure was repeated for the errors made by each individual subject, and the resulting distributions of $\chi^2$ obtained for the age groups were compared by a Mann-Whitney U test. The groups differed significantly ($p < 0{\cdot}01$ two-tailed), showing that associations of all kinds were more marked for the younger than for the older group.

B. *Distributions of errors between sub-groups of patterns.* Errors can be classified as "correct category errors" or "incorrect category errors." If a stimulus is a one-handed pattern a wrong response may be that appropriate to a different one-handed pattern (correct category error), or to one of the set of two-handed patterns (incorrect category error). Wrong responses to two-handed patterns can be classified in the same way.

# GROUPING OF STIMULI IN PATTERN RECOGNITION AS A FUNCTION OF AGE

BY

P. M. A. RABBITT

*From the Medical Research Council Applied Psychology Research Unit, Cambridge*

Two groups of subjects aged 20–40 and 40–60 years learnt identifying responses to a set of eight patterns. The groups did not differ in rate of learning, but an analysis of confusions between patterns revealed systematic differences in the extent to which rules underlying the construction of patterns were used in recognizing them as members of sub-sets before identifying responses were mastered.

## INTRODUCTION

It has been clear for some time that the simple association of a "stimulus" and a "response" is an inadequate description of the processes of adaptive selection which Bartlett (1932) has shown to be characteristic of human memory. A comparison between the rules of this selection and computer codes is a fashionable analogy (Cldfield, 1954 which has yielded interesting experiments (Ross, 1961).

Attneave (1955; 1957) and Fitts *et al.* (1956) have shown that subjects learning to identify patterns code the data they perceive so as to select cues which are discriminative for a set of patterns with which they are concerned, and to ignore cues which are misleading or redundant.

For any given ensemble of patterns there may be a single optimal set of cues having minimum redundancy, or a number of such sets representing equally efficient systems of classification. Sets of patterns may be constructed so as to offer hierarchical systems of classification, so that two cues may have to be considered in identifying a pattern, the first identifying a sub-set of which the pattern is a member, and the second identifying the pattern within this sub-set. A subject optimizing his coding system by learning such conditional cues may be said to be mastering the rules of classification of the set of patterns. Since relevant cues cannot be recognized, nor irrelevant cues ignored until the subject has some knowledge of the set of patterns as a whole, the progress of mastery of the rules of classification will be reflected in confusions made between patterns before they are identified completely.

A series of experiments by Rabbitt (1961) showed that older subjects are less efficient than young subjects when required to respond to categories of stimuli. Further, older subjects were found to be less able to ignore irrelevant information in perceptual tasks. The following experiment was undertaken to discover whether these results could be generalized for coding processes in pattern identification.

## MATERIAL

Eight patterns were used. Each consisted of two groups of four adjacent rectangles drawn in Indian ink on white pasteboard cards. Each rectangle measured 1½ in. × 4 in., and the two groups of rectangles were 2 in. apart. In each pattern two of these rectangles were inked in and the remaining six left unfilled. The patterns are shown symbolically below. X represents a filled rectangle and O an unfilled one.

*One handed patterns*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | O | O | O | O | X | O | O | X |
| 2 | O | O | O | O | O | X | O | X |
| 3 | X | O | X | O | O | O | O | O |
| 4 | O | X | O | X | O | O | O | O |

*Two handed patterns*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | O | O | X | O | X | O | O | O |
| 6 | O | O | O | X | O | X | O | O |
| 7 | X | O | O | O | O | X | O | O |
| 8 | O | O | X | O | O | O | X | O |

The eight patterns divide naturally into two sub-sets of four patterns each. In one sub-set both filled rectangles were in the same group of four (one-handed). In the other sub-set one of the filled rectangles was in each group of four (two-handed).

*Procedure*

Subjects were tested in groups of six, all members of a group being of the same age-range (20–40 or 40–60 years). The experimenter presented the patterns by holding up the cards one at a time for 5 sec. Intervals between successive presentations were of the order of 20 sec. and were occupied by subjects recording their responses on numbered protocols.

At the beginning of an experimental session subjects passively observed while the experimenter showed the eight patterns twice in random order and read aloud the responses assigned to them. The eight responses to be used, which were always "Bob," "George," "Arthur," "John," "Dick," "Peter," "Harry," and "Jim," were then copied down by the subjects. This was done to eliminate errors due to the use of unassigned responses, which would give no information about confusions between patterns. The eight responses were assigned to the patterns at random, assignments being varied between successive groups of subjects to eliminate persistent response-confusions from the summed data. When subjects had listed the required responses the patterns were presented singly, in random order, until each had been shown 13 times. After each presentation subjects recorded their responses, and the experimenter then read aloud the correct response.

At no time during the experiment was any comment passed which might indicate a possible scheme of classification for the patterns.

*Subjects*

Subjects were 58 postmen. Twenty-nine were aged 20–40 years (mean 27·4) and 29 were aged 40–60 years (mean 47·1).

## Results

1. *Rate of learning of correct identifications.* Error scores for each successive block of 26 trials were obtained for each group (see Table I) and compared by a Mann-Whitney U test. No significant differences were found. The groups can thus be equated for rate of learning, providing a criterion of comparability directly related to this task.

TABLE I

Error Scores for Successive Blocks of 26 Presentations

| | *1st Quarter* | *2nd Quarter* | *3rd Quarter* | *4th Quarter* |
|---|---|---|---|---|
| Youngs .. .. | 461 | 392 | 337 | 267 |
| Olds .. .. | 507 | 426 | 341 | 306 |

2. *Error distributions*

A. *Overall randomness of errors.* Errors made by each group of subjects were summed in a 7 × 8 error matrix (see Table II) and tested by $\chi^2$ to examine the hypothesis that they were randomly distributed. The 20–40 age group gave a $\chi^2$ of 259 and the 40–60 group a $\chi^2$ of 125, for 55 degrees of freedom showing that errors were not made at random by either group ($p < 0{\cdot}001$).

This procedure was repeated for the errors made by each individual subject, and the resulting distributions of $\chi^2$ obtained for the age groups were compared by a Mann-Whitney U test. The groups differed significantly ($p < 0{\cdot}01$ two-tailed), showing that associations of all kinds were more marked for the younger than for the older group.

B. *Distributions of errors between sub-groups of patterns.* Errors can be classified as "correct category errors" or "incorrect category errors." If a stimulus is a one-handed pattern a wrong response may be that appropriate to a different one-handed pattern (correct category error), or to one of the set of two-handed patterns (incorrect category error). Wrong responses to two-handed patterns can be classified in the same way.

## TABLE II

### ERROR MATRICES FOR YOUNG AND OLD GROUPS

Group I

Subjects aged 20–40 years

Pattern Presented

| Subject's Response | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | **229** | 48 | 23 | 17 | 12 | 12 | 18 | 18 |
| 2 | 45 | **198** | 19 | 26 | 30 | 26 | 18 | 15 |
| 3 | 23 | 40 | **207** | 43 | 14 | 18 | 16 | 16 |
| 4 | 20 | 23 | 39 | **222** | 16 | 19 | 21 | 17 |
| 5 | 16 | 28 | 23 | 22 | **207** | 30 | 28 | 23 |
| 6 | 22 | 13 | 20 | 15 | 33 | **176** | 56 | 42 |
| 7 | 22 | 17 | 36 | 28 | 40 | 36 | **198** | 24 |
| 8 | 24 | 20 | 23 | 23 | 51 | 40 | 50 | **145** |

Group II

Subjects aged 40–60 years

Pattern Presented

| Subject's Response | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | **185** | 35 | 18 | 20 | 25 | 31 | 35 | 28 |
| 2 | 43 | **199** | 12 | 17 | 27 | 27 | 26 | 26 |
| 3 | 20 | 15 | **209** | 36 | 19 | 18 | 25 | 35 |
| 4 | 24 | 12 | 34 | **172** | 37 | 44 | 23 | 31 |
| 5 | 25 | 21 | 21 | 33 | **187** | 17 | 39 | 34 |
| 6 | 27 | 35 | 17 | 31 | 30 | **174** | 38 | 25 |
| 7 | 29 | 29 | 25 | 25 | 41 | 41 | **151** | 36 |
| 8 | 32 | 28 | 33 | 33 | 30 | 27 | 35 | **159** |

Numbers of patterns refer to illustration in text.

On the assumption that the distribution of errors is random between correct and incorrect categories, the probability of an error occurring within the correct category by chance is 3/7 or 0·43. It was thus possible to use *t* tests to determine whether the means for the ratios of correct category errors to total errors obtained for each group differed from chance expectation.

These results, and the values of *t* obtained for each group are given in Table III. It can be seen that while the group of young subjects made significantly more correct category errors than would be expected by chance, errors made by the older group are random with respect to this classification.

It was also possible to test whether the difference between these ratios represented significant changes in performance with age using the ratio of Correct Errors/Total

### TABLE III

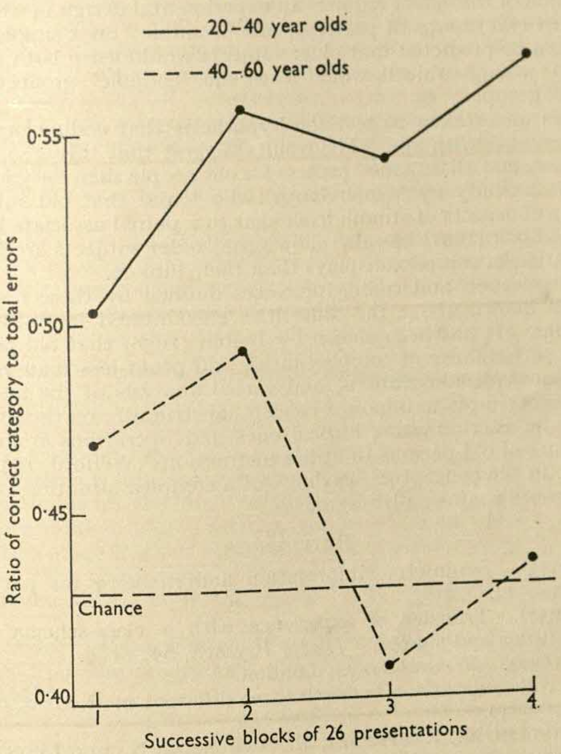
### Age and Type of Error

| *Group* | *Total errors* (*Max possible* = 3016) | *Correct category errors/ total errors* |
|---|---|---|
| 20–40 | 1457 (48·3 per cent.) | 0·56 ($t = 3·1$; $p < 0·01$) |
| 40–50 | 1580 (52·4 per cent.) | 0·43 ($t = 0·91$; $p > 0·25$) |

Errors obtained for each subject. A Mann-Whitney U test showed that the younger group made significantly more correct category errors than the older group ($p < 0·01$ two-tailed).

C. *Correct Category Errors as a function of learning.* Ratios of Correct Category Errors/Total Errors were obtained for each of four successive blocks of 26 trials comprising the 104 presentations given to each subject. These data are shown in Figure 1 below.

FIGURE 1



The incidence of correct category errors for successive quarters of the total number of presentations was tested against chance as described above. It was found that young subjects made more correct category confusions than would be expected by chance at all four stages ($p < 0·01$). The ratios obtained for older subjects did not differ from chance expectation for any quarter. Further, ratios obtained for the young subjects were significantly greater than those obtained for the old subjects ($p < 0·01$ Mann-Whitney two-tailed) for all except the first quarter.

The percentage of correct category confusions made by the younger group increases with practise. Ratios obtained for the last 52 trials are significantly greater than ratios obtained for the first 52 trials ($p < 0·01$ Wilcoxon two-tailed). This is not true of the older group, who make fewer of these confusions later in practise than during the first 52 presentations ($p < 0·05$ Wilcoxon two-tailed).

## DISCUSSION

Older subjects are found to be more random in their errors than younger subjects (Results: 2A), and not to use a particular system of classification employed by young subjects (Results: 2B and C).

In the context of previous work on pattern recognition (Attneave 1955; 1957; Fitts *et al.* 1956) this may be described as a situation in which subjects learn to ignore redundant cues guided by their growing knowledge of the constraints underlying the set of patterns as a whole. Older subjects seem either to be unable to perceive such constraints, or perhaps not to use their knowledge to guide their guessing during learning.

Since the younger subjects use what would appear to be a more efficient strategy, the fact that the old and young groups did not differ in rate of learning requires comment. Recognition of the correct sub-set raises the probability of guessing right by chance from 0·125 to 0·25, and young subjects made 14 per cent. more correct category errors than would be expected by chance. The utility of this in terms of an increase in correct responses is consequently very small in the present situation, but may be expected to rise sharply with the number of patterns to be learnt, and the number of sub-sets into which they may be grouped. It may be expected that discrepancies in efficiency between young and old subjects would become more obvious under these conditions. The further investigation of this effect requires an experimental design in which young and old subjects both learn two groups of patterns, one "codable" on a simple scheme, and one not. It might then be predicted that older subjects would learn both groups of patterns equally well, while young subjects would learn the "codable" group more readily than the "non-codable" group.

This study was undertaken to test the hypothesis that coding processes in pattern identification deteriorate with age. The results suggest that this is so, and that pattern identification is more an "all or none" process for old people than for young. This finding may be related to a study by Wimer (1959) who found that old subjects show more incidental learning of aspects of stimuli irrelevant to a paired associate learning task than young people. Rabbitt (1961) has also shown that older subjects are less able to ignore irrelevant information in complex displays than their juniors.

The failure of selection and coding processes implied by these results may also be related to what is known about the difficulties encountered in training old people in industrial situations. It has been shown by Belbin (1958) that old people are difficult to instruct in the performance of complex tasks, and profit less than young people from demonstrations. Such demonstrations, and verbal analysis of the task by instruction, may be regarded as attempts to impose a conceptual structure on the situation which will guide the trainee in discriminating critical cues and operations from incidental ones. The notorious failure of old persons to utilize instructions (Welford, 1958), may thus also represent a failure in the conceptual "coding" of a complex situation, and consequently, in the ability to ignore confusing irrelevancies.

## REFERENCES

ATTNEAVE, F. (1955). Symmetry, information and memory for patterns. *Amer. J. Psychol.*, **68**, 209–22.

ATTNEAVE, F. (1957). Transfer of experience with a class-schema to identification learning of patterns and shapes. *J. exp. Psychol.*, **54**, 81–8.

BARTLETT, F. C. (1932). *Remembering.* London.

BELBIN, E. (1958). A comparison between three different methods of operator training. *Ergonomics*, **1**, 39–40.

FITTS, P. M., WEINSTEIN, M., RAPPAPORT, M., ANDERSON, N., and LEONARD, J. A. (1956). Stimulus correlates of visual pattern recognition. *J. exp. Psychol.*, **51**, 1–11.

OLDFIELD, R. C. (1954). Memory mechanisms and the theory of Schemata. *Brit. J. Psychol.*, **45**, 14–23.

RABBITT, P. M. (1961). *Perceptual Discrimination and the Choice of Responses.* Cambridge Univ. Ph.D. Thesis.

ROSS, J. (1961). Human memory: A partial model and its implications for retroactive phenomena. *Princeton Univ. Educ. Testing Service Report.*

WELFORD, A. T. (1958). *Ageing and Human skill.* London.

WIMER, R. E. (1960). Age differences in incidental and intentional learning. *J. Geront.*, **15**, 79–82.

*Manuscript received 14th November, 1962.*

# CROSSMAN'S CONFUSION-FUNCTION AND MULTI-CHOICE DISCRIMINATION

BY

I. M. HUGHES*

*From the Clothing and Equipment Physiological Research Establishment, Farnborough*

In a recent paper Laming (1962) discusses afresh the relation between multi-choice reaction times and both the difficulty of discrimination and the entropy of the choice situation, a topic first discussed by Crossman (1953, 1955). To quote Laming, Crossman (1955) "examined the reaction time for several different levels of discriminability, and found the reaction times approximately linear with $C(S_1, S_2) = 1/[\log x_1 - \log x_2]$ which he called a "confusion-function." When this confusion-function was generalized to more than two choices, it seemed to give a complete explanation of reaction times, regardless of the increase in entropy."

However, Laming shows some doubt about Crossman's generalization of the "confusion-function," and for that reason restricted his experimental situation to the two choice situation. A little consideration shows that his doubt is well-founded.

The multi-choice situation discussed by Crossman (1955, to which all further references will be made), is that of "a sequence of signals in which n different alternatives $S_1, S_2 \ldots\ldots S_i \ldots\ldots S_n$ may occur; let them occur independently in the sequence with equal frequencies; further let each be recognizable only by its measure in a single dimension, the values being $x_1, x_2 \ldots\ldots x_i \ldots\ldots x_n$. We wish to measure the confusion for the set of n; that is, the average difficulty of deciding which of the Ss a given unknown Q is."

We should like to know the programme of comparisons between $x_q$ and the $x_i$ that leads to the identification of Q as one of the $S_i$. We can say that for any postulated programme, such as Crossman's, to be satisfactory, it must meet the following requirements.

(1) At any stage in the programme the $x_i$ chosen for comparison should not in general show any bias for or against the correct value of $x_q$, except as a result of information obtained from earlier stages; in particular, there should be no such bias on the first stage, in the absence of pre-presentation information. (Although it must be admitted that when the writer was a subject in an auditory experiment run by Dr. Crossman and Mr. R. L. Gregory in 1954, he responded disconcertingly often with the word that was about to be presented when he could not hear the word that had been.)

(2) If the programme is incomplete it should tell us most about the stages that make the largest contributions to the final response probabilities. In other words, if, say, stage A of the programme results in a tentative identification which is checked, but very rarely altered, during stage B, then we should want to know more about the comparisons involved in stage A than about those in stage B.

(3) The programme should be just as applicable to wrong identifications as to correct ones.

(4) If we can derive the distributions of errors and response times against signals and against each other from the programme, then they should agree at least roughly with experimental findings.

When we examine the programme put forward by Crossman we find that it fails to meet either requirement (1) or (2) and either (3) or (4).

He writes, "The operation we may imagine a subject's brain to perform is this. When a signal Q arrives, it is tentatively identified as one, $S_i$, which will in most cases be the correct one. The signal or its copy is then compared successively with every other possibility, $S_j$, and the answer 'different' is obtained for each after a time proportional to the confusion $C_{ij}$ between the two. Only after all these answers have been obtained will Q have been positively identified as $S_i$. Since the time taken in each comparison is proportional to $C_{ij}$, the time to identify $S_i$ will be"

$$t_i = A \sum_{j=1}^{n} C_{ij} = A.C(i) \quad \ldots \quad \ldots \quad \ldots \quad (i \neq j).$$

---

*Now at the Department of Building, Manchester College of Science and Technology.

Consider first requirement (3), that the programme should be equally applicable to errors. The expression for $t_i$ is made up of terms like $1/[\log x_k - \log x_i]$, $k \neq i$, when Q is tentatively and correctly identified as $S_i$. However if Q were tentatively and incorrectly identified as $S_k$, then the time to make the incorrect response $Q = S_k$ apparently includes the term $1/[\log x_i - \log x_q] = 1/[\log x_i - \log x_i] = \infty$. One might not always be correct but one would *never* make a mistake!

It may be that when Crossman writes of the signal *or its copy*—a term he uses nowhere else, and does not define—he means that one should substitute for $x_q$ not the correct value $x_i$, but the incorrect value, $x_k$. In that case, the third requirement is not violated—but the fourth one is, for according to the formula the subject will always take the same time to respond "$S_k$," whether the response is right or wrong, and if wrong, no matter what the correct answer is. Thus, consider the set of signal values {3, 7, 8, 9}. If 3 is presented and the subject says "7," then response time, $t \propto C(7) = 8{\cdot}76$, but if 3 is presented and subject says "9," then $t \propto C(9) = 9{\cdot}27$. On the other hand if either 7 or 9 is presented and the subject says "3," then $t \propto C(3) = 2{\cdot}16$.

Not only is the distribution of response times against errors unlikely but so also is the distribution of errors against signals, as far as one can gather, for Crossman appears to envisage the two-choice confusion-function as measuring the increase in response time against increase in difficulty with constant error rates. He writes for example, "The most satisfactory way of handling [errors] seemed to be to attempt experimentally to keep them constant on the average for all subjects"; again in the passage just quoted he writes, "the answer 'different' is obtained for each [comparison]" without mention of the possibility or the consequences of varying degrees of difference; and when he is testing the hypothesis that the brain is performing a statistical operation he makes no allowance, by changing the criterion or "confidence level," for differences in error rates. (It must be said that on these terms the rejection of the statistical hypothesis is hardly justified, since the points for all four subjects for the most difficult condition have been mis-plotted in Figure 4, and if they are plotted correctly the fit is not too unsatisfactory. Mis-plottings indeed occur throughout the paper: the abscissa for the easiest condition in Figure 3; the highest point in Figure 5; the regression lines in Figures 6 and 7; the abscissa for C(4, 12) in Figure 7; and all the points for subjects C and D in Figures 2, 3, 4 and 6.)

If error rates *are* to be constant in the two-choice case, then this would seem to imply that a presented signal is equally likely to be mistaken for any other in the set, e.g. in the set already considered, {3, 7, 8, 9}, if a 9 is presented and the subject makes a mistake, then he is as likely to say "3" as he is to say "8." If the error-rates are not to be constant, then we could reasonably ask for at least a hint as to how (and why) they should vary with difficulty.

Now consider requirement (1)—that the programme should not initially favour the correct answer—and requirement (2)—that the stages contributing most to the response probabilities should receive the most explicit treatment. If Crossman meant what he wrote about the unknown signal to be taken seriously, that when it arrives "it is tentatively identified as one, $S_1$, which will in most cases be the correct one," then we can quite fairly ask how long this identification takes, whether the time is influenced by the number of signals in the set, and by the values of the other signals, and so on; in other words we re-open the whole question of multi-choice discriminability. The proposed programme has failed to meet requirement (2), and has evaded the question.

On the other hand it may be that the tentative identification is not meant to be closely considered, and has been postulated largely to overcome the serious difficulty that arises if the correct signal is not removed from the sequential checking process. This difficulty is similar to that already noticed in the case of wrong responses, and arises from the fact that, according to the confusion-function, the time taken to compare Q and $S_1$ is proportional to $1/[\log x_1 - \log x_q]$, and when $Q = S_1$, this time equals infinity; so that if the correct signal value *does* enter the sequence of comparisons, and the confusion-function is applied to all the comparisons, then all response times will be infinitely long. But if the tentative identification has been proposed simply as a solution to this problem, then the cure is as damaging as the ailment, since it then violates requirement (1). The proposed programme begs the question.

All is not yet lost, for we can save the confusion-function in the two-choice case. For instance suppose the two comparisons, $x_q$ with $x_1$, and $x_q$ with $x_2$, to take place simultaneously, and as before, suppose $Q = S_2$. Then, after a time, $t \propto 1/[\log x_1 - \log x_q]$ the subject will have the answer "different" for the first comparison, and since the knowledge that $Q \neq S_i$ is sufficient to tell him that $Q = S_2$, the infinite comparison can be terminated.

A programme like this of simultaneous comparisons could quite feasibly be applied to the multi-choice case, and it would preserve the argument of Crossman's programme that identification is achieved by rejecting all except one of the possibilities; but in order to relate multi-choice times to values of the two-choice confusion-function we should have to have a number of fresh assumptions about the effects of the size of set, etc. In other words, this programme would result in quite a different generalization from that proposed by Crossman.

Yet another approach is possible. Suppose that $x_q$ is compared, not with the two signal values in the two-choice set, but with some average of them—for convenience, the geometric mean—then perhaps the time for a comparison will be related to $1/[\log (x_1 x_2)^{\frac{1}{2}} - \log x_q] = 1/\frac{1}{2}[\log x_1 - \log x_2] = 2C(S_1, S_2)$.

Now this approach will generalize to the multi-choice case. Suppose $x_1 > x_2 > \ldots > x_n$, and that $x_q$ is compared successively with the geometric means of adjacent signal values. The general term will be $C' (k, k + 1; i) = 1/[\log(x_k x_{k+1})^{\frac{1}{2}} - \log x_i]$ when $Q = S_i$, and the confusion for the ith signal will be given by

$$C'_1 = \sum_{k=1}^{n-1} C' (k, k+1; i)$$

As we saw, $C' = 2C$ for the two-choice case. However, all multi-choice values of $C'$ will be slightly less than $2C$, since $1/[\log (x_k x_{k+1})^{\frac{1}{2}} - \log x_i] = 1/\frac{1}{2}[C \log x_k - \log x_i) + (\log x_{k+1} - \log x_i)] < 2C(k, i)$ and $< 2C(k+1, i)$ unless $i = k+1$ or $i = k$. Thus $C'$ (4, 5, 6, 8, 10) is about 13·2, whereas $2C$ (4, 5, 6, 8, 10) is 15·8. We should therefore expect to see multi-choice times lying below the regression line for two-choice times when plotted against C. There is little or no sign of this in Crossman's results, suggesting that this new derivation does not fit the facts. And in any case, even this limited correspondence between old and new depends upon an implausibility, that the subject compares $x_q$ with all the means. Yet once he knows that $x_q < (x_k x_{k-1})^{\frac{1}{2}}$ and $> (x_k x_{k+1})^{\frac{1}{2}}$, he can identify Q as $S_k$, and there is no point in making further comparisons. If the subject makes his comparisons according to the optimum, halving, strategy, then the average number of comparisons required to make an identification (if none of the comparisons gives the wrong answer) will be close to $\log_2 n$—the entropy! If the subject starts with the smallest standard, say, and works steadily up, then the average number of comparisons will be just over $\frac{1}{2}n$. If the subject makes the comparisons in random order, the average number required will be roughly $\frac{2}{3}n$. And although response-time will not be proportional to the number of comparisons involved, yet when we consider the allowances that would have to be made for the occasions when one or more of the comparisons produced the wrong answer, and for the different times that would be taken over the different sequences of comparisons that could precede the crucial two, and when we consider also that it is arguable—as we shall do in another paper—that the whole approach represented by the confusion-function, of trying to find a "tight" relation between response time and difficulty of discrimination, is mistaken, when we consider all these, then it would hardly be worthwhile to calculate the actual response times we might expect from these different strategies—but at least it is clear that for all of them multi-choice response times must fall even further below the regression line of two-choice times against C than when all the comparisons are made.

Crossman writes, "We tentatively conclude that the confusion-function fits multi-choice tasks of up to five alternatives of a one-dimensional signal." He presents the results for six subjects in a table, and for one of them in a graph in addition. For this subject the fit is a good one. When one plots the results for the other five subjects, the fits are much less satisfactory; not everyone would conclude—even tentatively—that the generalization was justified.

In plotting two-choice data (e.g. Fig. 6) Crossman subtracts movement-time from sorting-time to give a choice-time, although he does not need to. In plotting multi-choice results he does not subtract movement-times, although it is at least possible that this time varies with the number of choices.

There are a number of discrepancies between values of the confusion-function given or derived from Table IV and those given in Table V.

These are shown in the table below:

| | C(4,5)* | C(3,4,5) | C(3,4,5,6) | C(3,4,5,6,7) |
|---|---|---|---|---|
| From Table IV .. | 3·13 | 4·60 | 6·71 | 8·83 |
| From Table V .. | 3·08 | 4·85 | 6·90 | 8·95 |

* Correct value = 3·11

It is quite possible that the confusion-function does give a rough indication of the difficulty of a situation, of the kind that could be useful to the human engineer, but so would any number of functions, many of them easier to calculate. For instance, if we assume $x_1 > x_2 > x_3 > \ldots\ldots > x_n$, then $\sum_{k=1}^{n-1} \frac{\sum_{i=1}^{n} x_i}{x_k - x_{k+1}}$ would probably prove just as useful; and even the sum of the signal values divided by their range might be sufficient for most purposes. Both these formulae reduce in the two-choice case to $\frac{x_1 + x_2}{x_1 - x_2}$, which (Welford, 1960) closely approximates the confusion-function.

When discussing the relation of the multi-choice results to Information Theory findings, Crossman says, "In theory discriminability (or confusion) and information content should be inter-related at least for signals in one dimension when an upper limit is set to signal size." But if discriminability is based on the ratios of signal magnitudes and not on their absolute values, then this statement is, if not incorrect, certainly incomplete, and would be true only if there were also a lower limit, greater than nothing, set to signal size.

## References

Crossman, E. R. F. W. (1953). Entropy and choice time. The effect of frequency unbalance on choice response time. *Quart. J. exp. Psychol.*, **5**, 41–51.

Crossman, E. R. F. W. (1955). The measurement of discriminability. *Quart. J. exp. Psychol.*, **7**, 176–95.

Laming, D. R. J. (1962). A statistical test of a prediction from information theory in a card-sorting situation. *Quart, J. exp. Psychol.*, **14**, 38–48.

Welford, A. T. (1960). The measurement of sensory-motor performance: survey and reappraisal of twelve years' progress. *Ergonomics*, **3**, 189–230.

*Manuscript received 25th June, 1962.*

# REPLY TO DR. I. M. HUGHES

BY

E. R. F. W. CROSSMAN

*From the Institute of Experimental Psychology, Oxford*

Hughes' closely-reasoned letter has raised too many points for me to discuss in a short note, so I will consider only the status of my confusion-function and whether comparisons are simultaneous or successive in multi-choice discrimination. My views on the latter point changed substantially not long after writing the 1955 paper, both for the reasons that Hughes has now stated so cogently, and also on considering fresh evidence, so I shall only be able to answer Hughes' criticism of my earlier position obliquely by stating my present one.

First, though, Hughes has found a number of discrepancies between my tabulated results and the graphs derived from them. On rechecking I agree that—(*a*) In Figure 4 the four points for the 10/12 condition are about 5 mm. too low. (*b*) In Figure 3 the four points for the 1/10 condition are about 4 mm. too far to the right. (*c*) In Figure 5 the highest point is almost 1 mm. too near the axis. (*d*) In Figures 6 and 7 the regression lines should be slightly steeper. (*e*) In Figure 7 the points for the 4/12 condition are about 2 mm. too near the axis. However, (*f*) the points for subjects C and D in Figures 3, 4 and 6 are otherwise correctly plotted. They were obtained by subtracting the overall mean movement-time (0·553 sec.), not the individual ones (0·551, 0·554, 0·535, and 0·568 sec.) from the mean response-times, which may have misled Hughes. I can only apologize for errors (*a*) to (*e*), caused by my unskilful drafting; of these, however, (*b*), (*c*) and (*d*) do not affect my argument; (*e*) goes slightly against it, and only (*a*) is potentially misleading. Here, however, my rejection of the hypothesis in question was based on the other end of the curve, which is unaffected by the correction. I agree that the hypothesis cannot confidently be rejected because of Figure 4 alone, but Figure 6 was definitely more promising and, as Hammerton (1959) and Laming (1962) have since pointed out, Equation (4) is only one of several possible outcomes of a statistical theory.

The discrepant values for C(4,5), etc., in Table V were derived from an early version of Table IV drawn up with a slide-rule; on recalculating it for publication (working to three decimal places and rounding off to two) I should also have corrected the entries in Table V, and I am grateful to Hughes for drawing attention to my oversight. Here again the errors, though inexcusable, do not affect the conclusions drawn from Table V, such as they are.

Turning now to theoretical issues, my paper dealt with two distinct questions:

(1) What is the law relating choice-time to signal values in a two-choice one-dimensional discrimination?

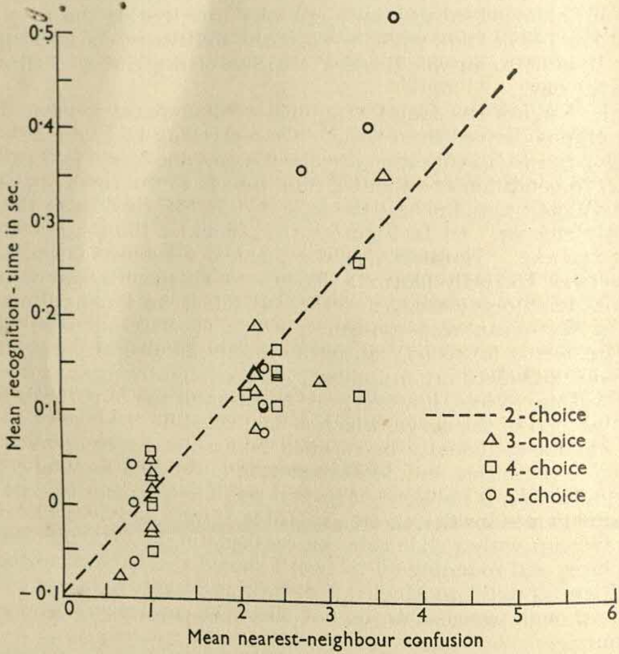
(2) Can such a law be generalized to multi-choice one-dimensional discrimination?

The further question (3) Can it also be generalized to multi-dimensional discrimination? was also in my mind and conditioned the conclusions but I did not raise it explicitly in the paper.

Hughes apparently accepts Equation (6) as a valid answer to question (1), but he claims that other expressions fit equally well, and Welford (1960) has indeed demonstrated that the expression $(x_1 + x_2)/(x_1 - x_2)$ is an "approximately equivalent and rather more convenient equation." I entirely agree. But of course the purpose of this (and any) basic research is more than finding a formula of use to human engineers; we want to discover the mode of action of the system being studied, in this case the human discriminative system. The purpose of curve-fitting is to test predictions from given hypotheses or structural models, and since neither Welford nor Hughes have shown that my third model gives a *bad* fit to my data, nor have they proposed a better rational hypothesis (as distinct from a formula), there is no case for me to answer. Welford points out that the data of Birren and Botwinick (1955) do not fit; but in their experiment, as in Henmon's (1906), the discriminanda were lines presented side by side, and the discriminant (i.e. the *single* variable actually used to make the choice) cannot be identified with confidence. It could have been either length alone or some geometrically constructed combination of both. So these experiments cannot be directly compared to mine. In a like situation with *successive* presentation a difference of 1·32 per cent. would be well below threshold and the choice-time would thus be effectively infinite.

Hughes also objects that my hypothesis does not predict the frequency or type of errors. But it was never meant to. Although, paradoxically, with card-sorting subjects must be permitted to make errors if they are to work at top speed, by slowing down slightly they could in principle perform perfectly, and my hypothesis was framed for the latter case. Certainly a complete (but necessarily more complex) theory would deal with errors (see e.g. Stone, 1960; Laming, 1962) and my result would then represent the limiting case as error tended to zero.

FIGURE 1



Recognition-time and confusion in multi-choice sorting tasks (data from Crossman, 1955, Table V).

Each point is the mean of three trials by one subject with movement-time and calculated response selection-time subtracted. The ordinate is the mean value of Confusion (op. cit., Table IV) between each signal in a set and its nearest neighbour. The two-choice results are shown as a regression-line (t = 0·11C − 0·09 sec.).

Turning to Question (2), Hughes points out that to assume an initial "tentative" identification in a multi-choice situation is begging the question. I plead guilty. The postulate was introduced to make a successive-comparison model possible without having *Q* compared to itself, and I was uneasily aware that a wrong initial identification produced nonsense. In effect I was supposing that the time was all spent in a secondary checking process rather than in primary recognition. But the alternative simultaneous-comparison model seemed to have the worse fault that it did not predict increased times for larger sets, as found empirically (Table V). On later consideration, though, it has come to seem most unlikely that all of the increase could in any case be accounted for by differences in discriminability alone, as was proposed. In Table V, for instance, by subtracting the last row from the next to last we obtain the choice-times for two-choice tasks of very low confusion (0/1, etc.); they do *not* approach zero, as they should according to the view proposed in the paper. They lie between 155 and 261 millisec., corresponding to information-rates between 3·8 and 6·5 bits/sec., in rough agreement with Hick's (1952) and Crossman's (1953) results. This suggests that Hick's law (the logarithmic dependence of choice-time on number of alternatives) applies essentially to low-confusion tasks, and hence one may infer that the choice-time is spent in selecting the *response* rather than in recognizing the signal. Now as recognition of signal must precede selection of response, the extra recognition-time needed in high-confusion tasks must be added to the response selection-time given by Hick's law. To obtain true recognition-times experimentally, then, one should subtract the time taken for a control task with the same number of alternatives but very low confusion, from that for the experimental task.

My experiment IV included no such control tasks, but one can extrapolate from the two-choice times using Hick's law and thus correct the choice-times given in Table V. The outcome is shown in Figure 1, where it can be seen that time still increases with confusion. These data do not warrant more precise analysis, but the even distribution of points for different numbers of alternatives now seems to favour a simultaneous-comparison model. Much more information could be obtained by recording choice-times for individual responses in one-dimensional multi-choice tasks and studying their variation under changes of ensemble.

Further support for a simultaneous-comparison model can be found in the results of experiments such as those of Davis, Moray and Treisman (1961) where choice-time was little affected by size of ensemble. Since the same signals—spoken digits and letters—can be used in other tasks showing large variations of choice-time, the latter cannot be ascribed to the recognition process itself which is presumably the same in both cases. Also Broadbent and Gregory (1962) found very small differences between choice-tasks where the same response was given to one or to two different signals, and Fraisse and Blancheteau (1962) have found only 2·5 millisec. difference in visual duration-threshold between 2 and 15 printed nonsense-syllables, corresponding to an information-rate well over 1,000 bits per sec. Thus recognition-time seems to be little affected by size of ensemble. The successive-comparison model would predict a large effect, so these experiments tell against it.

In short, then, I accept Hughes' strictures on my successive-comparison model, and I withdraw my tentative conclusion that Hick's law can be explained without reference to information theory. Instead I wish to draw a clear distinction between time for discrimination (or recognition), and time for selection of response, the former being some function of the *signal* set, and the latter proportional to *response* entropy. We need more empirical data before carrying the theory further, and I hope Hughes will report his results without delay.

## References

Birren, J. E., and Botwinick, J. (1955). Speed of response as a function of perceptual difficulty and age. *J. Gerontol.*, **10**, 433–6.

Broadbent, D. E., and Gregory, M. (1962). Human response to classes of stimuli. *Nature*, **193**, 1315–6.

Davis, R., Moray, N. and Treisman, A. (1961). Imitative responses and the rate of gain of information. *Quart. J. exp. Psychol.*, **13**, 78–89.

Fraisse, P., and Blancheteau, M. (1962). The influence of the number of alternatives on the perceptual recognition threshold. *Quart. J. exp. Psychol.*, **14**, 52–5.

Hammerton, M. (1959). A mathematical model for perception and a theoretical confusion function. *Nature*, **184**, 1668–9.

Henmon, V. A. C. (1906). The time of perception as a measure of differences in sensation. *Arch. Philos. Psychol. Sci. Meth.*, **8**, 73.

Hick, W. E. (1952). On the rate of gain of information. *Quart. J. exp. Psychol.*, **4**, 11–26.

Lamimg, D. R. J. (1962). A statistical test of a prediction from information theory in a card-sorting situation. *Quart. J. exp. Psychol.*, **14** 38–48.

Stone, M. (1960). Models for reaction-time. *Psychometrika*, **25**, 251–60.

Welford, A. T. (1960). The measurement of sensory-motor performance: survey and reappraisal of twelve years' progress. *Ergonomics*, **3**, 189–230.

*Manuscript received 7th January, 1963.*

# APPARATUS

## AN APPARATUS FOR PRESENTING CONTINUOUSLY VARYING BINOCULARLY DISPARATE VIEWS OF AN APPARENTLY SINGLE OBJECT

BY

WILLIAM GROMAN

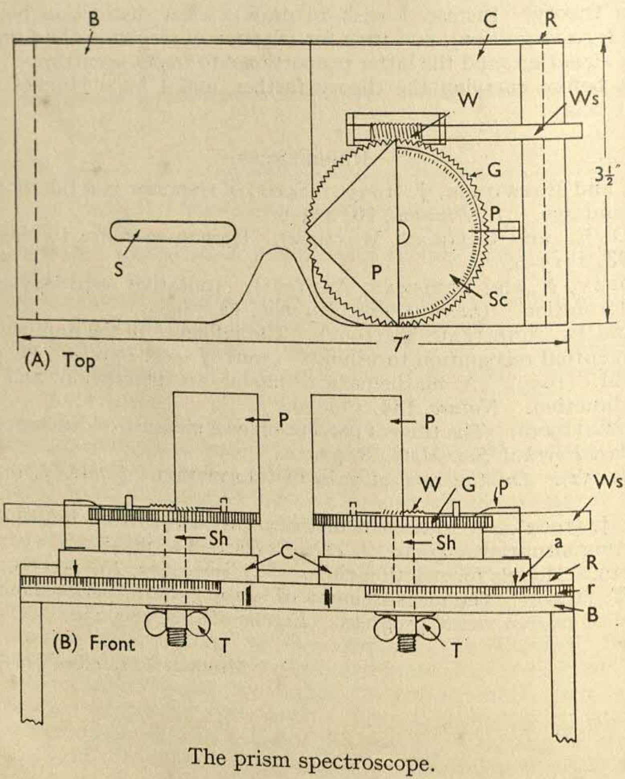
*From Department of Psychology, Duke University*

The perceptual effects of viewing continuously varying binocularly disparate views of a stimulus object under conditions in which convergence and accommodation are kept relatively constant and normal with respect to the target object have not been investigated. Information regarding the effects of such stimulation should have relevance for problems of form and pattern perception, object constancy, and the perception of depth. The apparatus to be described makes it possible to produce the stimulus conditions necessary for the study of this class of effects.

### *Apparatus*

The apparatus consists of two major parts: the prism stereoscope through which an observer looks, and the stimulus carrier, consisting of two rotating circular discs, so coupled that they may be rotated simultaneously in opposite (or the same) direction(s). (Such a stereoscope presents the right image to the left eye and left image to the right eye. However, since the stimulus objects are identical pseudoscopic effects do not occur.)

FIGURE 1



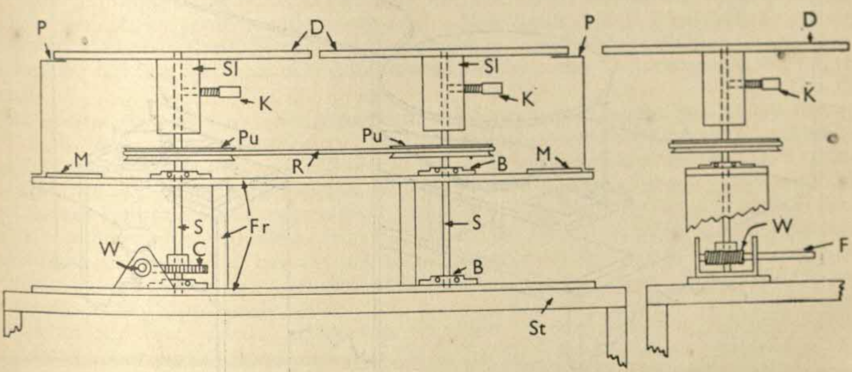
The prism spectroscope.

On each disc one of two identical objects, the stimulus objects, may be so placed that the image of each is presented only to one eye. The rest of the apparatus serves the purposes of controlling speed of rotation, size of the field of view, and the nature of the background against which the stimulus objects are seen.

As shown in Figure 1, the *prism stereoscope* consists of two 45° optical glass prisms, base, $2\frac{1}{4}$ in., height, $1\frac{1}{2}$ in., mounted so that the distance between them may be varied to be optimal for a given interpupillary distance. Each prism carrier *C*, shifts to left or right in its slot *S* independently, and is locked by the thumbscrew *T* for this adjustment. Thus each prism is individually adjustable for distance from centre of bridge of the observer's nose. Settings are read on the straight scale, *r*, by means of inscribed indicator, *a*.

In addition, each prism may be independently rotated around the centre of its base, by means of the worm gear shaft, *Ws*, so that the lateral overlap of the visual field, for each eye may be adjusted for satisfactory stereopsy with regard to any visual idiosyncrasy of the observer. Appropriate scales and indicators, as shown, are mounted in order to record settings.

FIGURE 2



The stimulus carrier.

Figure 2 shows the mounting and arrangement for rotation of the two discs, at the top centre of each of which the two identical stimulus objects are mounted. The discs are 9 in. in diameter and are mounted with $\frac{1}{8}$ in. between circumferences. A variable speed motor coupled to *F*, the shaft of the worm gear, rotates the shafts, *S*, in opposite directions when the rubber belt, *R*, is fitted with a one-half twist on pulleys, *Pu*. Speeds of rotation can vary from 20 per min. to one in two min. This wide range was chosen to make possible a wider exploration of conditions. On the underside of each disc a circular protractor scale is mounted. By means of indicator, *P*, and mirror, *M*, settings of the discs may be made or read. For easy reading, mirror images of the appropriate numbers were drawn on the protractor scales.

Figure 3 shows how the instrument is assembled, and indicates position of the supplementary parts. A screen, *Sc*, conceals the stimulus carriers from the observer, who rests his chin in the chin cup, *C*, and through an adjustable aperture, *V*, sees (after adjustment) only the apparently single stimulus object and some portion of the discs and background. A mounting board, *B*, permits different backgrounds to be used.

The light source, not shown, is mounted horizontally above the discs, and consists of two 18 in. fluorescent tubes, parallel to the screen, and lighting the fronts of the two identical stimulus objects from above. Screen *Sc*, extends vertically to a height concealing the light source from the observer.

Relevant distance measurements used in assembling the instrument as constructed follow:

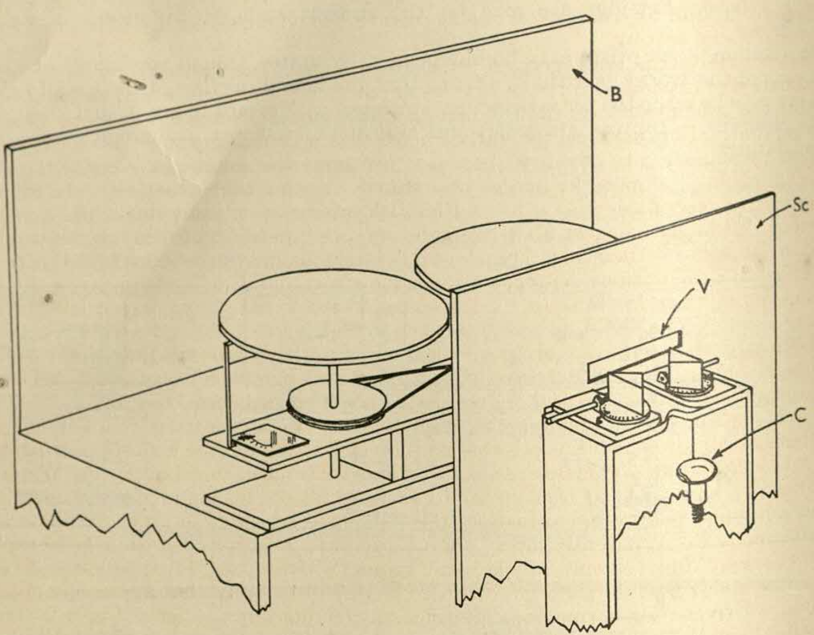
(1) The vertical centre of the prisms are $17\frac{1}{4}$ in. above the table top.

(2) The tops of the discs are $15\frac{5}{8}$ in. above the table top.

(3) Distance from centre of prism bases to centre of discs is 17 in. (The stimulus appears to the observer to be approximately 18 in. from his eyes. Convergence and accommodation can be comfortably maintained, therefore, for relatively long periods of time, e.g. 1 hr.)

(4) Mounted on a 31 in. high table top, this results in a convenient height for seated observation.

FIGURE 3



The assembled apparatus.

*Discussion*

The apparatus is quite versatile. It affords the opportunity of studying a number of phenomena during motion as well as at rest, utilizing the same stimulus material. The variety of stimulus objects which may be used is practically unlimited.

For encouragement and assistance in preparing this note, thanks are extended to Dr. K. E. Zener, Duke University, and to Mr. R. L. Gregory, Cambridge University.

*Manuscript received 23rd October, 1962.*

# BOOK REVIEWS

*Aggression: a Social Psychological Analysis.* By Leonard Berkowitz. London: McGraw-Hill Book Co. Inc. 1962. Pp. xv + 361. 62s.

This book is an attempt to draw together diverse studies of aggression into a single theoretical system. To this end the author synthesizes a wide range of material on man and animals, and invokes the techniques of experimental psychology, personality theory, social psychology and other disciplines. Although the problems considered range from child-rearing practices to international warfare, the diverse material is successfully synthesized around an analysis of the aggressive behaviour shown by individual human beings in a social setting.

Unfortunately, the book gets off to a sticky start. The first chapter is devoted to an evaluation of instinct theories of aggression, meaning effectively Freud and McDougall. Whether the conclusions are right or wrong, the argument is marred at many points by a *non sequitur*. For instance, the author holds that male sex hormone "does not provide any direct stimulation to hostility" because "learning experiences can override the effects of the androgen." Again, he quotes two studies showing that experience can influence the aggressive responses of cats to rats and of ants to other individuals and concludes that they "strongly suggest that animals are not predetermined to aggress against particular objects by their genetic make-up." Also, having justly criticized Freud's use of teleological explanations, he repeats the same error, confusing consequence with cause, in discussing fighting for rank in a dominance hierarchy and "non-competitive" fighting in mice.

The author then sets out his theoretical framework—a modification of the Dollard, Doob, N. E. Miller, Mowrer and Sears 1939 frustration/aggression hypothesis. Aggression is the consequence of frustration of "instigated response sequences" or "drives." (These two terms are said to be synonymous, implying "an ongoing activity.") To make this hypothesis as comprehensive as possible he is obliged to stretch the concept of frustration until it becomes almost valueless: thus he follows McDougall in ascribing the aggressive retaliation of a man who has been struck to a frustration of "his impulse of self assertion." Further, although poking fun at instinct theorists for the number of urges that have been called instinctive ("Apparently, every time many of these writers felt an urge to explain some behaviour they would invoke an 'instinct' "), the author apparently has no reservations about postulating a new drive to be frustrated every time aggression must be explained. Drives "for harmonious and pleasant relationships," "for self-preservation," "for self-esteem," "to do well," "for self-enhancement," even by implication "for religious certainty" are provided *ad lib.*

There are two main amendments to the Dollard thesis. First "the instigation to aggression" produced by frustration is identified as anger, which is supposedly definable by physiological indices. Anger heightens the probability of aggression, though its actual occurrence may depend on the presence of suitable cues, past experience and so on. Second, one way in which past experience influences the outcome of frustration is through the individual's interpretation of the situation. "Expectancy" is introduced here, though in an everyday sense and without rigid definition. Here is a brave attempt to marry different approaches to the study of behaviour, and even different levels of analysis: one's confidence would be increased if one could be sure that the difficulties had been recognized.

The frustration/aggression hypothesis has thus been extended into real life situations, but without the rigour used by N. E. Miller in his attempts to "liberalize" S-R concepts. But when he is dealing with the everyday world, or trying to trap it in the laboratory with experiments on his students, Dr. Berkowitz comes into his own. Here he becomes hard-headed, buttressing his conclusions with caveats, emphasizing the deceptiveness of correlational findings, insisting that the findings of laboratory experiments may be misleading in real life, and showing a readiness for an all-round, comprehensive and yet critical approach lacking in the earlier chapters. The limits of the frustration-aggression formula are noted and the importance of a multidisciplinary approach underlined.

The result is a critical contribution to some of today's most important problems—racialism, aggressive personalities, child rearing, crime. His analytical approach to hostility catharsis explodes many misconceptions, and his measured assessment of the role of violence on mass media should be read by television promoters. It is to be hoped that the contents of these chapters will be assimilated not only by psychologists.

R. A. HINDE.

*Cumulative Record.* By B. F. Skinner. London: Methuen. 1962. Pp. x + 426. 50s.

The work of B. F. Skinner is still little known in this country despite the fact that *The Behavior of Organisms*, his major scientific contribution, was published as long as 25 years ago. Now, however, both its theoretical interest, and its wide range of application in such diverse fields as education, pharmacology, sociology, psychotherapy, and indeed in every branch of the behavioural sciences, is bringing it to the attention of psychologists. There is no better way to become acquainted with it than through this book.

The Skinnerians' own behaviour is partly responsible for the resistance shown to their work. They make no attempt to relate their work to that of their more conventional colleagues, they publish for the most part in their own journal (*Journal of the Experimental Analysis of Behavior*) in a terminology which abounds with technicalities, and many of them display an arrogant impatience with views outside their frame of reference. The major heresies of the movement are the use of hypothetical models, the practice of statistically analysing data from large groups of subjects, and the retention of outmoded non-automatic apparatus such as the jumping-stand, the activity wheel and the maze.

On the other hand the main articles of faith are simply that behaviour is an orderly process which is best studied by the detailed observation of individual subjects, that a suitable basic datum is frequency of response, and that a system of laws of behaviour can be built up by using simplified situations where variables can be isolated and where the organism comes fully under the control of the experimenter.

*Cumulative Record* is a selection of Skinner's papers, slightly enlarged from that published in the United States 4 years ago and adequately reflects his three main interests.

His first major contribution is, of course, in the field of animal behaviour, and 15 papers in the present collection give the reader good coverage of this area. The best of these is the one entitled, *The Experimental Analysis of Behavior*, which describes the standard techniques, and gives a representative selection of achievements. Further papers deal with the gradual shaping of responses by successive approximation, anxiety as measured by the technique of conditioned suppression, superstitious behaviour, and the interesting war-time attempt to use pigeons to guide missiles.

The most controversial paper in this section is that called, *Are Theories of Learning Necessary?* which has often been interpreted as an argument against all types of theorizing. Here Skinner is rather confusing since he uses the term *theory* in a particularly restrictive sense to mean only those theories which attempt to explain the data in terms of events at some other level of observation. Thus he would include the kinetic theory of gases, and the synaptic theory of reflex action, but not Kepler's theory of elliptical orbits or Newton's Laws, since the latter, although transcending particular facts, are no more than a generalization of relations between observable variables. That he is in favour of this second type of theorizing is made quite clear in a later paper on *Current Trends in Experimental Psychology*, which is a more sophisticated version of the same argument. Logically he is on shaky ground here since the first type of theory has proved its worth in Physics time and time again, and will no doubt eventually do so in Psychology, but as a statement of priorities at this stage of the science it has much to recommend it. Certainly most of the current theories have a phlogiston smell about them, and a great deal of research stimulated by theory is irrelevant not because the theory has been proved wrong but simply because it has gone out of fashion.

The success of operant conditioning in controlling the behaviour of experimental animals leads naturally to its application in human affairs and Skinner's other two major interests lie in the technology of education, and the control of human behaviour toward social ends.

His system of programmed instruction is a direct application of principles from the animal laboratory. The material to be learned is broken up into small steps and presented in such a form that the student must compose an appropriate response for which he is immediately reinforced. In general it is found that the knowledge that he has been correct is sufficient reinforcement, and the work is so carefully graded, each step being dependent on previous ones, that subjects should make few if any errors. Skinner believes that automatic programming is necessary to get optimal results and has designed teaching machines for this purpose. However he himself has published a programme for psychology students in book form, and the use of the term *machine* with its computer overtones is misleading since the essential element in the system is the programme not the method of presentation. The success of the method depends very much on the skill of the programmer, but it shows promise of relieving teachers of the drudgery of imparting facts, giving them more time for a personal role.

If the behaviour of animals can be changed, why cannot men be taught to be happy, skilful, well behaved, and productive? Skinner feels that the difficulties here are political rather than scientific. But philosophical issues are also raised, particularly regarding the concept of freedom. Effective control will be by positive reinforcement rather than by the present punitive methods, and men will be free to do what they wish, with their wishes rather than their freedom controlled. To those who shrink from these implications Skinner retorts that we are already controlled by the world in which we live, and that the choice is whether we will be " . . . controlled by accident, tyrants, or by ourselves in effective cultural design." As with most utopia builders over-simplifications abound, but in a nuclear age, cultural engineering is surely a matter of urgency which few psychologists show signs of appreciating.

The rest of the book is disappointing. The application of operant techniques to psychiatric problems shows signs of promise but as yet little achievement. It is not easy to transfer from the laboratory to real life situations, where social reinforcers dominate, but one feels that basic problems will not be solved by making psychotics press levers.

A few theoretical papers on such topics as the distinction between respondent and operant conditioning will be familiar to many and probably of interest only to new-comers to the field, and a section on literary and verbal behaviour has hardly enough substance to warrant inclusion. Indeed the book would have been greatly improved if cut to half its size. Of course one doesn't have to read it all and it is well worth discarding the gristle to savour the meat.

J. STEINER.

*Psychology Through Experiment.* Edited by G. Humphrey. London: Methuen. 1963. Pp. 307. 30s.

This book tries explicitly to kill two birds with one stone, even though the birds are certainly not of a feather and therefore present a difficult target. A common fate of such an ambitious aim is of course to miss both birds, and it is therefore a relief to find that although one escapes, the other is struck a firm, even if only glancing blow. What species of birds? They appear in the two introductions, one for humanists the other for psychologists. We learn that the book attempts to give non-psychologists a grasp of what psychologists are doing and why, and at the same time to provide instruction in psychological experiments that are suitable for classwork. The seven contributors cover the following topics: sensory processes, perception, remembering, learning, motor performance, thinking, and experiments with animals.

The chapter on sensory processes makes three main points. First, we need a model of how things might work before investigating how they do work. Second, a particularly fruitful model is information theory, whereby we consider the flow and handling of information rather than, as in classical physiology, the transfer and conduction of energy. Third, when we apply this model to a study of sensory processes it emerges that a number of apparently quite different sensory discriminations, such as flicker, acuity, and colour vision, can be reduced to a common basis, namely intensity discrimination. While many may prefer a different kind of model for their approach to these problems, few will deny that in this chapter it is used to great advantage in quite ingenious experiments.

The chapters on remembering and learning are, I think, the most lucid in the book, perhaps because they restrict themselves so severely, the former to a comparison of recall and recognition and the effect on these of set and interpolated activity, the latter to a consideration of the concept of association and the possible nature of the "link." Each contributor has space to develop his theme and illustrate it with closely-related, well controlled experiments.

The chapters on perception, motor performance, and thinking are, by contrast, much more loosely organized and less fluent. The experiments they include are individually interesting but they do not relate clearly to any theory or model of perception, skill, or thought. For instance, the chapter on perception has only three introductory pages about the theoretical basis of this vastly complex topic, followed by 49 pages of experiment and references. Will the reader appreciate the relationship that is suggested between the phi phenomenon and the "Köhler-Wallach Satiation hypothesis, and the general theory of cortical isomorphism" when the latter are not in the least explained? This kind of mistake occurs throughout the book.

The last chapter is called "Experiments on Animals." To commission it was a bold decision and to write it must have been extremely difficult. But neither of these considerations can excuse its shortcomings. Despite the title, the only animal concerned is the laboratory rat, and the experiments relate entirely to two theories of the behaviour of rats in mazes, the author of one theory being also the author of the chapter. This might not matter if either theory was adequately explained; but neither is, with the result that the experiments are often mystifying. It could well be that this author is so familiar with his own theory and that of his rival that, like most of us, he assumes far too much knowledge on the part of the novice to whom he is giving a brief resumé of his own work.

In summary, this book succeeds in describing in considerable detail experiments which are most appropriate for groups of students in a well equipped laboratory. If the students already have appreciable background knowledge or the time to follow up the abundant references, they will find much of interest in it. So too will instructors who have to design practical classes. However, without pursuing the references it will be difficult for the novice to understand the experiments or why they are being done, for with few exceptions little is said in the book about broad theoretical issues. This handicaps the non-psychologist and he is the bird that gets away.

A. Cowey.

*Behaviour and Awareness.* Edited by C. W. Eriksen. Durham, N.C.: Duke University Press and London: Cambridge University Press. 1962. Pp. vi + 158. 28s.

This symposium contains interesting papers by six authors, five of whom are concerned to develop empirical methods for studying "awareness," whilst the sixth, Verplanck, argues that the word is quite dispensable. Since "awareness" is a systematically ambiguous term, it is perhaps regrettable that several of the authors do not clearly indicate to which types of awareness their methods and conclusions are intended to apply. The discussions of learning-without-awareness often fail to specify of what aspect of the learning situation the individual is or is not aware. For instance, each of the five authors who considers learning-without-awareness (the exception being Hilgard) at some point deals with learning without awareness of a response-reinforcement contingency. But several of these authors discuss learning without awareness of cues or stimuli, or learning without awareness of any reinforcement. It is often difficult to determine what is intended in a particular context.

Indeed learning without awareness of a response-reinforcement contingency turns out to be the central topic of this book. This finding comes as something of a disappointment after the title has indicated a treatment of Behaviour. Both Eriksen and Kimble in fact note the distinction between learning and performance, but do not pursue the notion that individuals can "perform" without awareness. Yet, although the emphasis is on learning, no mention is made of insight-learning, which to some appears as the paradigm of learning without awareness. The inclusion of Hilgard's paper on awareness of stimuli in perception does, however, provide some justification for the comprehensive title.

The obstacle to any experimental attack on awareness is, of course, that of providing an operational definition of the term. The contributors to this symposium agree that a reliance on verbal reports is likely to constitute the most profitable spearhead. Results of studies employing non-verbal measures of awareness (as reported, for instance by both Eriksen and Hilgard) are extremely difficult to interpret. Dulany and Speilberger employ verbal measures and both reach the conclusion that verbal conditioning occurs only when individuals are aware of the response-reinforcement contingency (or hold a "correlated hypothesis") and have a positive attitude towards the reinforcement.

Yet there are opposing conclusions in the recent literature. This points to a difficulty attached to the use of verbal measures of awareness. We can be fairly sure that, if a verbal measure indicates awareness, the individual was in fact aware. But if the measure fails to indicate awareness, we cannot draw any firm conclusion; either the individual was in fact not aware, or the measuring instrument is not sufficiently precise to reveal awareness. The measures described by several authors in this symposium are, however, likely to prove more effective than earlier methods of assessing awareness through verbal reports.

Every one of the contributors reports results of some hitherto unpublished research, and another useful feature of this book is the relatively detailed description and discussion of work previously published in a variety of journals and texts. It is, therefore, unfortunate that no index is provided. This symposium is valuable in that it suggests empirical approaches to old and new problems, but the problems still remain largely unsolved.

Peter B. Warr.

in disturbed personalities. Under the "Physiochemical Regulation of Growth and Behaviour" there is surprisingly little on the hormones, when compared with the space allotted to the effects of drugs on behaviour. The reviewer would also like to have seen more mention of attempts at "synthetic" models of the reacting organism, though Bindra's and Selye's work is discussed.

The work will probably appeal most to students trying to view behaviour theory dispassionately. To analytically-minded persons the book as a whole may appear unsophisticated, but others may find this refreshing. It is possible that, despite its title as a textbook, it may have some circulation amongst experienced workers looking for a convenient summary on complex behaviour.

D. J. MURRAY.

*The Assessment of Pain in Man and Animals.* Edited by C. A. Keele and Robert Smith. London. U.F.A.W. (The Universities Federation for Animal Welfare). Distributing agents: E. & S. Livingstone. 1963. Pp. xi + 323. 30s.

This book consists of the papers given at an international symposium held under the auspices of U.F.A.W. in London during 1961. Symposia are notoriously difficult to discuss in a reasonably short space, and this one is no exception, since there are 27 separate papers describing and discussing recent work in different fields.

Lord Brain contributes the Presidential Address, and the concluding lecture on "Avoidance of Pain in the Laboratory" is given by that veteran warrior of U.F.A.W., Major C. W. Hume. The symposiasts are a number of distinguished anatomists, neurologists, pharmacologists, physiologists, and veterinary surgeons, from America, Britain, France, Germany, Sweden and Switzerland. Each reviews recent work in his own general field, and there are many excellent diagrams and photographs. The peripheral and central nervous mechanisms of pain are discussed in detail, as are the stimuli which produce pain and the modes of action of some of the drugs which reduce it.

The book is well produced. It has refreshingly little about the duller forms of psycho-physics or about the philosophical aspects of the mind-body problem; but it contains a wealth of empirical information.

G. C. GRINDLEY.

*A Study in Behaviour.* By S. A. Barnett. London: Methuen. 1963. Pp. xvi + 288. 45s.

Mr. Barnett's book invites discussion on two levels. In the first place it is a convenient review of behavioural work on the rat. The author moves with considerable versatility between ethology and psychology, and into physiology, with chapters covering Exploration, Feeding Behaviour, Rat Societies, Fixed Patterns, Learning, Drive, and the Brain and Behaviour. Material is presented selectively and critically, with no attempt at exhaustive coverage; there is, however, an excellent bibliography. Psychologists are most likely to profit from the earlier chapters which will remind them that rats live private lives in nests and burrows, as well as public lives in mazes and Skinner boxes. It is, however, instructive to read a brief and well-proportioned survey of learning theory by someone who is not a learning theorist.

In the second place the book is explicitly concerned with the logic and methods of behavioural enquiry; it is sub-titled "Principles of Ethology and Behavioural Physiology, Displayed Mainly in the Rat." Mr. Barnett writes as critic rather than evangelist, and no single approach to behaviour is strongly favoured. He is refreshingly impatient of imprecision ("However, all this is very vague: what do we mean by activation?") and shows us clearly the reduced logical circumstances of terms like "drive" and "instinct." He is scrupulously careful to define his own terms, to the extent of adding one appendix on the nature of definition, and a second giving "stipulative definitions" of terms as he has used them. There are, however, occasional lapses of rigour; the discussion of Hebb's views on perceptual learning could usefully have been more critical, and the need to describe some behaviour as "insightful" is perhaps too readily accepted.

The writing is fresh and lucid, and the book is unusually well produced, with black and white plates supplemented by most attractive line drawings.

M. H. SHELDON.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XVI 1964 Part 3

# THE EFFECT OF STEADY FIXATION ON THE JUDGEMENT OF RELATIVE DEPTH

BY

IAN P. HOWARD and WILLIAM B. TEMPLETON
*From the University of Durham, England*

Subjects were asked to judge the relative depth of test edges after various conditions of inspection: (1) when the edges were equidistant; (2) when no edges were present; (3) when one edge was 1, 2, or 4 in. in front of the other.

Condition 3 produced a significant shift of the subjective equidistance point towards the inspection positions. The results fit in with an explanation in terms of a Gibson-type adaptation process.

There is no evidence of stereoscopic fatigue or binocular image "inversion." A figural after-effect in depth, if present, is small compared to the adaptation effect. An explanation in terms of cyclotorsion or other eye movements seems impossible. A central shift in the stereoscopic disparity mechanism could account for the results, but so also could a monocular scale shift produced by the "anchor" effect of the inspection edges.

In a study of reversible perspective in three-dimensional objects (Howard, 1961), it was hypothesized that steady fixation of objects at different relative depths would lead to a disruption of relative depth perception. A search of the literature revealed very few papers on the effect of long fixation on relative depth perception.

Köhler and Emery (1947) in their work on figural after-effects in the third dimension of visual space were the first to mention the possible effects of steady fixation on relative depth perception. In their first experiment they presented an inspection (I) object in the frontal plane. A test object was subsequently shown in the same location turned through 25° or 45° backwards or 25° or 45° forwards. Each time, the T or test object appeared as if displaced from the frontal position of the I-object. In a second experiment they showed that after fixation of a concave I-figure, a flat T-figure appeared convex. Monocular viewing was said to reduce these after-effects substantially. In a third experiment they showed that T-objects always recede from I-objects. Thus T-objects behind the position of a preceding I-object were pushed backwards and T-objects in front were pushe[illegible]ward. These opposed effects on the T-object, it was argued, exclude the possibility that after-images cause the T-objects to appear dimmer, and to recede because of this. Up to a point, the extent of the effect of the I-figure was found to increase as its distance from the T-figure increased. This distance paradox was said to be analogous to the one found for two-dimensional figural after-effects.

Köhler and Emery (1947) discussed the possibility that the three-dimensional after-effects may be reduceable to after-effects in two dimensions. They pointed out that a test figure, having a different position in depth to an I-figure, will project images to each eye at least one of which must have a slightly different retinal position to the corresponding I-figure. This should cause the T-image or images to retreat from the positions occupied by the I-figures and these shifts could be interpreted by the subject as a change in the binocular disparity and hence the three-dimensional position of the T-object. In outlining this argument Köhler and Emery admitted its logical validity but claimed to have refuted it by an experiment. In this they presented, by means of a stereoscope, the left and right eye views of the I-figure alternately. This was said to destroy the three-dimensional appearance of the I-figure; but to leave a two-dimensional pattern which would reveal any purely two-dimensional, ordinary figural after-effect which this pattern of stimulation may have been producing. Köhler and Emery did not find any after-effect after this procedure and concluded that the occurrence of three-dimensional after-effects depends on the three-dimensional appearance of the pattern.

Köhler and Emery did not report their procedure in sufficient detail to show whether they were successful in eliminating the three-dimensional appearance of the I-figure; it is our experience that stereoscopic vision may persist down to four alternations a second. Even if they did eliminate it, they should still have got a two-dimensional after-effect on the basis of their own argument. Three subjects only were used in their experiments, and there has been no independent confirmation of their results.

Bergman and Gibson (1959) claimed to have demonstrated an after-effect of fixating a surface slanted in the third-dimension, which was not a figural after-effect. The stimulus was a textured surface, slanting at various angles either backwards or forwards. This was viewed through a tube so that no contours were visible. After 4 min. inspection, the subject was asked to set a similar textured surface to the subjective vertical. They found that the subjective vertical was displaced towards the I-figure. They suggested that this effect could not be due to figural after-effects because there were no contours in the I-figure. This is not a convincing argument, for the textured surface must have involved a density gradient in a fine pattern and there is no reason why a Köhler-type satiation process should not occur differentially over such a surface.

Bergman and Gibson found that the effect was nearly as strong when the I-figure was not steadily fixated. They used this result to support the theory that the adaptation effect is not specific to the locality of the retina stimulated. This conclusion seems ill-founded; the extent of eye movements was not stated. A better test of whether the effect is locus specific would have been to expose the I-figure in one retinal locality and the T-figure in another.

Ogle and Weil (1958) studied the effect of duration of the stimulus on stereoscopic acuity. They found a four-fold decrease in stereoscopic acuity associated with a decrease in duration of exposure from 1 sec. to 1/124th sec. They explained their results in terms of the number and extent of eye movements during exposure.

Lit (1959) studied the effect of fixation conditions on stereoscopic acuity using the Howard-Dolman apparatus. He had three conditions of fixation. Condition 1 involved steady fixation of the movable comparison rod, condition 2 involved steady fixation of the immovable standard rod, and in condition 3 the subjects were allowed to fixate either at will. Condition 2 produced the lowest precision and condition 1 the best. In all conditions there was a movable comparison rod; the subjects never fixated a constant arrangement of rods. The results of Lit's experiment are therefore not directly related to the present investigation, but it is of interest to note that in his condition 2 where the subjects most nearly approximated to constant fixation of a fixed pattern, the stereoscopic acuity was worst.

Ditchburn and Pritchard (1960) studied binocular rivalry with two stabilized retinal images. Their findings do not relate to the problem of stereoscopic acuity and are of no concern here, but the technique is potentially useful for investigating the effects of constant fixation on depth discrimination.

Künnapas (1960) investigated the psychophysical functions relating objective to subjective distance. He concluded his paper with the statement, "It is conceivable that, in analogy to adaptation to light intensity, temperature, etc., an adaptation of the subjective range to the stimulus range may change the exponent" (i.e. the exponent of the psychophysical function). Künnapas gave no experimental evidence for this statement, but it does foreshadow our own results.

## EXPERIMENT I

Constant fixation of a three-dimensional pattern may have several effects on stereoscopic vision. Four possibilities are listed below.

*Hypothesis A: Stereoscopic fatigue.* The continued fixation of a pattern having depth leads to a deterioration of stereoscopic vision for that particular pattern in that particular locality, or a general stereoscopic fatigue of the whole field. In a task demanding an equality setting of distant rods, stereoscopic fatigue would cause an increase in the variance of the settings but no shift of the P.S.E.

*Hypothesis B: Stereoscopic inversion.* After long fixation of a three-dimensional pattern, the central mechanism of vision confuses the inputs from the two eyes. This confusion of the inputs leads to an inversion of the binocular disparities, such as occurs in a pseudoscope. In the absence of strong cues to the contrary, things would appear to be inside-out.

Asher (private communication) has claimed that normal stereopsis may be produced by alternating the two stereoscopically disparate views of a scene to the same eye in rapid succession. This finding suggests that the C.N.S. is indifferent to the ocular origin of disparity information, and if true would make our hypothesis more plausible.

Hypothesis B, if true, would lead to the same results as stereoscopic fatigue. Both mechanisms could be localized or general in their effects on the test stimulus, and in either case would lead to an increase in the region of uncertainty, that is, a decrease in the slope of the psychometric function.

*Hypothesis C: Normalization to equidistance.* After fixating a three-dimensional pattern there is a shift of the norm of equidistance in depth such that the pattern comes to look flat. A corollary is that the relative distances of other three-dimensional patterns, occupying a similar retinal locality, will be judged in terms of the shifted norm of equidistance. There will be a general displacement of all relative distances in a direction opposite to the depth difference in the pattern originally inspected. This would be an adaptation or normalization effect, analogous to Gibson's normalization to the vertical (Gibson and Radner, 1937).

*Hypothesis D: Three dimensional figural after-effect.* This is based on Köhler's theory and states that test objects behind the position of a previously inspected object are apparently displaced backwards, and test objects in front are displaced forwards.

The following experiment goes some way towards testing these hypotheses.

*Apparatus.* There were two target patterns, an inspection pattern (I) and a test pattern (T). Each consisted of black cards arranged as in Figure 1. In each pair, the standard card remained stationary while the variable could be moved so as to be in front of, or behind the standard card. The lateral distance between the cards was constant; the cards were rigid, vertical and parallel to each other.

The I and T cards were positioned in planes at right angles to each other and at 45° to a semi-reflecting mirror placed between them. The I and T cards were thus exactly superimposed when viewed from the subject's position, 9 ft. in front of the cards. Each standard card had a small hole near to its edge and these holes, which were also superimposed, served as fixation points.

An opal glass screen illuminated from the back was placed behind each pair of stimulus cards, which thus formed black-white vertical contours. The total appearance was that of a white rectangle, the left edge of which was the standard and the right edge the variable.

Each opal screen formed the front side of a 15 in. cubical box, on the back inner surface of which were mounted six, 60 watt, filament striplights. The surface illumination of the screen was 40 foot-candles. The two screens were matched for brightness in the following way; half of each was covered with black card so that the experimenter, sitting in the position of the subject, saw the two remaining halves side by side. Bulbs were removed from one of the light boxes until a match was obtained. A black screen was matched to the shade of black of the stimulus cards so that no contours were visible to the subject except the two inside edges of the stimulus cards.

Each variable card had a pointer attached to it, which showed the position of the variable on a scale marked off into five positions each 2 in. apart. Figure 1 shows the arrangement of the five positions. A memory-drum exposure device presented to the experimenter a random series of the five position numbers, one at a time. The experimenter set the T card according to the number while the subject was viewing the I pattern.

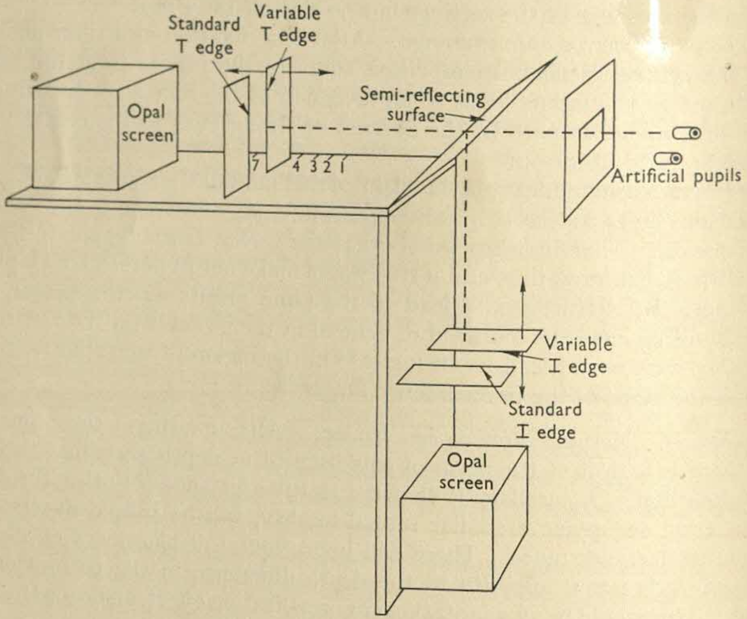


FIGURE 1

Apparatus used to demonstrate an adaptation effect in the judgement of relative depth.

Electronic timers controlled the duration of exposure of the I and T patterns and the movement of the memory-drum device. The subject signalled his response to the experimenter by pressing a key backwards or forwards after each exposure of the T pattern.

*Procedure.* The subject was seated 9 ft. away from the target cards, with his head firmly clamped. Artificial pupils were used, the right-hand one being fixed so that the right eye of the subject was in line with the variable edges. The instructions to the subject were as follows. "You will see a white rectangle; keep your gaze centred on the small white spot on the left of the rectangle. The left-hand edge of the rectangle will be stationary; the right-hand edge will be moved so as to be in front of or behind the left edge. Indicate whether you think it is in front or behind by pushing this switch in the same direction. You will be given only 1 sec. to make this judgement, there will then be a period of 10 sec. in which you will see the rectangle or just a spot of light. Do nothing except fixate during this period. A warning click will be given just before each 1-sec. test period, there will be 50 tests alternating with 10-sec. intervals. The initial inspection period will be 2 min. Always respond 'in front' or 'behind'."

Each subject was given a practice session and then a session for each of four experimental conditions. The sessions were well spaced, and their order varied for the different subjects. The experimental conditions were as follows.

*The equals condition.* Here the I-pattern was the white rectangle with equidistant vertical edges. A 2-min. inspection period preceded the 50, 1-sec., test exposures interspersed with inspection periods each lasting 10 sec.

*The disparity conditions.* These were similar to the "equals" condition, except that the right-hand edge of the white rectangle was 2 in. in front of the left-hand edge in one condition and 4 in. in front in the other, during the inspection period.

*The control condition.* Here the I-pattern was a small fixation light only.

The test periods were brief so as not to disrupt the state induced by the I-pattern. The 10 sec. reinforcements of the I-pattern were given to keep the effect "topped up." Each of the five positions of the variable edge of the test rectangle occurred 10 times in random order, making 50 readings in all. Each session occupied about 12 min.

*Results.* The raw data consisted of the number of responses in which the variable edge was judged "forward" of the standard edge. The total number of "forward" responses made by each subject in each condition was calculated, and the means of the five totals are shown in Table I for each condition.

TABLE I

RESULTS FROM EXPERIMENT I

| | *Conditions* | | | |
|---|---|---|---|---|
| | *Control* | *Equals* | *2 in. forward* | *4 in. forward* |
| Mean number of forward responses | 25·6 | 25·8 | 12·8 | 15·6 |
| $\sigma$ m. .. .. .. .. .. | 3·61 | 1·80 | 1·96 | 2·36 |
| P.S.E. .. .. .. .. .. | 0·08 | −0·05 | 3·08 | 2·18 |
| $\sigma$ mdn. .. .. .. .. | 0·78 | 0·37 | 0·51 | 0·44 |

TABLE II

ANALYSIS OF VARIANCE OF THE FORWARD RESPONSES FROM EXPERIMENT I

| *Source* | *d.f.* | *Sums of squares* | *Estimated variance* | *F* | *p* |
|---|---|---|---|---|---|
| Conditions .. .. .. | 3 | 140 | 46·7 | 16·2 | <0·001 |
| Stimulus categories .. .. | 4 | 125 | 156 | 54·2 | <0·001 |
| Subjects .. .. .. | 4 | 58 | 14·5 | 5·03 | <0·01 |
| Conditions × categories .. | 12 | 50 | 4·17 | 1·44 | N.S. |
| Conditions × subjects .. | 12 | 46 | 3·83 | 1·32 | N.S. |
| Categories × subjects .. | 16 | 125 | 7·81 | 2·71 | <0·01 |
| Residual .. .. .. | 48 | 138 | 2·88 | | |
| Total .. .. .. .. | 99 | 1,182 | | | |

An inspection of Table I suggests that fixation of a particular pattern with the variable edge in a forward position leads to a reduction in the number of forward responses, that is, to an apparent backward displacement of the variable stimulus in the test pattern. The estimated P.S.E's for each condition are also set out in Table I together with their standard errors. The results are in inches; positive quantities represent "forward" positions, that is, the variable edge appearing closer than the standard edge. The P.S.E.'s were obtained by averaging the 50 per cent. points (weighted two), the 75 per cent. points and the 25 per cent. points of the psychometric functions shown in Figure 2. An analysis of variance was done on the totals of "forward" responses and the results are set out in Table II.

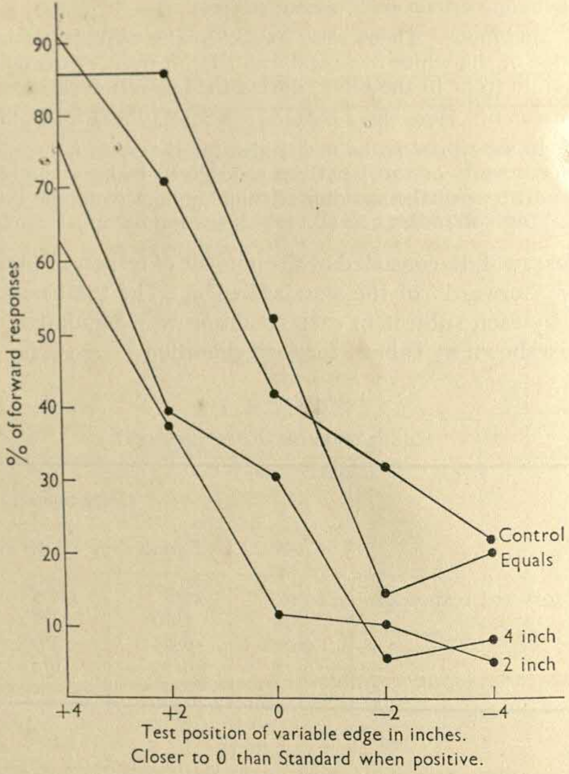


FIGURE 2

Psychometric functions for the four conditions of Experiment I.

Tested against a standard error based on the residual variance from the analysis, the mean difference between the "equals" and 4-in. conditions yields $t = 4{\cdot}16$, which for 48 degrees of freedom has $p < 0{\cdot}001$. The mean difference between the 2-in. and 4-in. conditions gives $t = 1{\cdot}42$, which is not significant. Hence the significant main effect of conditions is due to a real difference between two groups of conditions, equals and control on the one hand and 2-in. and 4-in. on the other.

## EXPERIMENT II

The second experiment was designed to check that the shift in P.S.E. was dependent on the inspection periods and not merely an effect of successive contrast.

*Procedure*. There was no initial inspection period and the periods between judgements were shortened from 10 to 5 sec. A second group of five subjects was used. In view of the level of discrimination of the subjects in Experiment I, it was decided to shorten the stimulus category intervals to $1\frac{1}{2}$ in. In view of the results of Experiment I, the control condition was omitted and the 4-in. inspection pattern was replaced by a 1-in. condition. Otherwise the design was identical with that of Experiment I. The analysis shown in Table III, is again of totals of forward responses.

*Results.* The results of Experiment II are set out in Table III. The mean difference between the equals and 2-in. condition yields $t = 3·7$, which for 32 degrees of freedom has $p < 0·001$. Hence the significant main effect of conditions is due to a real difference between the equals condition on the one hand and the 1-in. and 2-in. conditions on the other.

TABLE III

RESULTS FROM EXPERIMENT II

| | *Conditions* | | |
|---|---|---|---|
| | *Equals* | *1 in. forward* | *2 in. forward* |
| Mean number of forward responses .. | 28·8 | 21·6 | 21·4 |
| $\sigma$ m. .. .. .. .. .. | 1·88 | 1·36 | 1·96 |
| P.S.E. .. .. .. .. .. | −0·5 | 0·59 | 0·54 |
| $\sigma$ mdn. .. .. .. .. .. | 0·26 | 0·18 | 0·32 |

TABLE IV

ANALYSIS OF THE VARIANCE OF THE FORWARD RESPONSES FROM EXPERIMENT II

| *Source* | *d.f.* | *Sums of squares* | *Estimated variance* | *F* | *p* |
|---|---|---|---|---|---|
| Conditions .. .. .. | 2 | 35 | 17·5 | 9·35 | <0·001 |
| Stimulus categories .. .. | 4 | 945 | 236 | 126 | <0·001 |
| Subjects .. .. .. | 4 | 30 | 7·5 | 4·01 | <0·05 |
| Conditions × categories .. | 8 | 29 | 3·6 | 1·92 | N.S. |
| Conditions × subjects .. | 8 | 8 | 1 | 0·53 | N.S. |
| Categories × subjects .. | 16 | 32 | 2 | 1·06 | N.S. |
| Residual .. .. .. | 32 | 60 | 1·87 | | |
| Total .. .. .. .. | 74 | 1,139 | | | |

The shift of P.S.E. between the equals condition and the disparity conditions is considerably smaller than in Experiment I (see Fig. 3). Subtracting the mean P.S.E. in the equals condition from that in the 2-in. condition gives an effect of 3·13 in. in Experiment I, and 1·04 in. in Experiment II. In terms of frequencies of forward responses the difference between the equals condition means in the two experiments is not significant whereas the difference between the 2-in. means gives $t = 4·09$, which for 80 degrees of freedom has $p < 0·001$.

## DISCUSSION

Thus we conclude that when the variable edge of the inspection pattern is closer to the subject than the standard edge, there occurs a shift in the P.S.E. towards the subject, i.e. variable edges of test figures appear farther away. There is no evidence that the size of this effect depends on the degree of separation of the inspected edges, 1, 2 and 4 in. being equally effective. But the process has been shown to be time dependent, since the addition of an initial 2-min. inspection period together with a doubling of the inter-judgement inspection period produced a significant increase in the size of the effect.

The results will now be considered in relation to the four hypotheses which the experiment was designed to test.

*Hypotheses A and B: Stereoscopic fatigue and inversion.* According to these hypotheses, the variance in the control condition, when the stimulus edges were not present in the inspection period, should be lower than the variance in the other conditions. The standard errors of the means in Table I reveal that the opposite of this prediction is true. The hypotheses are therefore contraindicated. These hypotheses could not account for the shift in the P.S.E. which occurred in the disparity conditions.

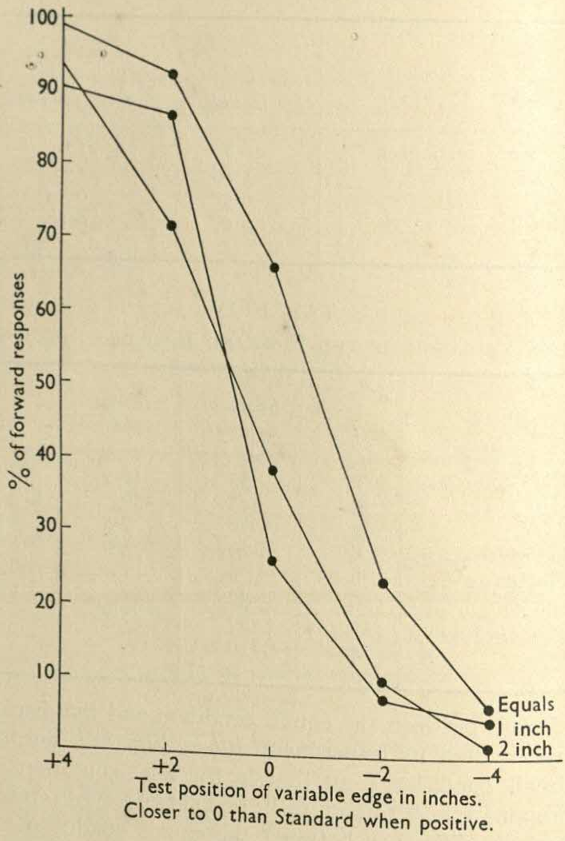


FIGURE 3
Psychometric functions for the three conditions of Experiment II.

It may therefore be concluded that no pronounced stereoscopic fatigue or inversion was present in the experiments.

There is the possibility that fatigue, or the tendency to inversion, affects only the test stimulus which corresponds to the inspection stimulus. If this were true, the response variability for those test stimuli would be greater than in the control condition. The results reveal no such tendency and therefore the presence of localized fatigue is not indicated.

*Hypothesis C: Normalization to equidistance.* From the analysis, the mean difference of forward responses between the control and both the 2-in. and the 4-in. conditions was significant, that is, there occurred a significant shift of the P.S.E. in a direction towards the inspection figure. Figures 2 and 3 show the psychometric functions from which it can be seen that the change in the number of forward responses was in the same direction for each of the stimulus categories. This implies a shift of the whole subjective scale of relative distance in the same direction.

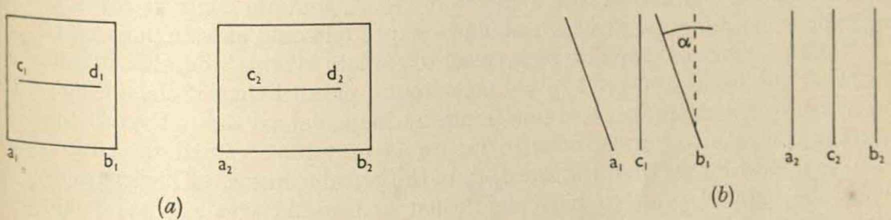
These results strongly support the hypothesis that there is an adaptation to the "norm" of equidistance, and that this process is formally analogous to those normalization processes described by Gibson and Radner.

*Hypothesis D: Three-dimensional figural after-effect.* There is no strong evidence that variable edges suffer an apparent displacement away from the position of inspected edges as would be predicted from Köhler and Emery's satiation theory. But such effects may be masked by the tendency of scales to remain all of a piece with reference to the P.S.E., even when the latter undergoes a shift. There is, however, an indication of such a figural after-effect in data from Experiment I, shown in Figure 2. The equals condition produced disproportionately large and small numbers of "forward" responses to the +2 in. and −2 in. test stimuli respectively. These are the test stimuli adjacent to the position of the inspection figure. However, this discrepancy was not sufficiently large to give a significant interaction between conditions and categories.

Similarly, the variations in the slope of the curves for Experiment II (Fig. 3) though not significant, are in the expected direction. The category interval 3 in. to 1·5 in. includes the position of the 2-in. inspection figure and the 2-in. curve is the steepest of the three. The interval 1·5 in. to zero includes the position of the 1-in. inspection figure, and here the 1-in. curve is the steepest. There is no inspection figure contained in the interval zero to −1·5 in. but the equals curve, with inspection figure at zero, is the steepest.

The main result of the present experiments may thus be summarized in two statements: (1) there is a tendency for non-equidistant objects to appear equidistant; (2) this tendency outlives the presence of the inspection stimulus and reveals itself as a shift in the whole scale of relative distance, in subsequently exposed test stimuli.

Statement 1 is confirmed in the work of others.



(*a*) (*b*)

FIGURES 4*a* AND 4*b*

Patterns used by Werner to demonstrate binocular depth contrast.

Werner (1938) exposed the two patterns of Figure 4*a* in a stereoscope. The frame did not appear to slant much in depth despite the considerable disparity of the half-images. He assumed that the primary correspondence was altered. As a consequence, the two equal length lines $c_1$–$d_1$ and $c_2$–$d_2$ were judged according to the new correspondence, and appeared to slant in the opposite direction to the frame. This so-called binocular depth contrast could be thought of as the simultaneous counterpart of the successive adaptation effect reported here.

Werner reported a similar effect with the stereograms shown in Figure 4*b*; there, two sets of lines were fused and the disparities in the lines $a_1$, $b_1$ and $a_2$, $b_2$ were said to induce a contrast tilting of the lines $c_1$ and $c_2$.

Ogle (1946) attempted to explain this last effect in terms of cyclofusional eye movements which he said would be induced by the disparity of the outer lines. These cyclofusional movements would, he claimed, restore the correspondence between the outer lines and consequently introduce a disparity between the inner ones. This type of explanation is only possible for contrast effects produced by slant about a horizontal axis. In our experiments the relative depth changes are equivalent to a slant about a vertical axis and neither eye torsions nor other types of eye movement could "cancel" the retinal disparity.

Gogel (1956) demonstrated an "equidistance tendency." This is a tendency for all parts of an object and for different objects to be seen in the same frontal-parallel plane when depth cues are weak.

Harker (1962), in a study of the role of cyclotorsion in binocular contrast, concluded that there are, as well as cyclofusion, "perceptual factors" which determine the apparent displacement of objects to the frontal-parallel position. One of the perceptual factors he suggested is the equidistance tendency.

While all this evidence supports an interpretation of our results in terms of a central shift in the scale of binocular disparity, or in Werner's terms, a shift in stereoscopic correspondence, an explanation in terms of monocular factors cannot be ruled out.

In our experiment, the separation between the images of the standard and variable stimulus edges was constant in the right eye. We need only consider, therefore, the changes in the left eye. In the experiment the subject is shown a series of 50 random settings of the edges. The possibility exists that during this test series he builds up a subjective scale of monocular image separations of the edges. He could use such a scale as an indication of relative depth. In terms of adaptation theory, his neutral or equidistance point in this subjective scale would tend to centre about the mean of the stimulus series. In the control and equals condition this would correspond to objective equidistance, but in the 2-in. and 4-in. inspection conditions the inspection pattern would produce a retinal separation of the images of the edges in the right eye which would be to one side of the mean of the separations of the test series. A typical "anchor" situation would therefore result, and the shift in the subjective mean or equidistance point which would be expected could explain our results.

Monocular factors may also be invoked to explain Werner's depth contrast. The fact that the frame appeared to be more frontal-parallel than it should, may have been due to the suppression of one frame by binocular rivalry. Even if this were not the case, the narrower frame in Figure 4*a* may have caused the enclosed line $c_1$–$d_1$ to appear longer than the line $c_2$–$d_2$ in the broader frame. This length contrast would be induced prior to binocular fusion and would give rise to an apparent binocular disparity. The effect would not depend on the frame being seen in depth and would not therefore be a depth contrast effect.

In Figure 4*b*, the tendency to see as vertical, lines which are tilted away about a horizontal axis, which Ogle explained in terms of cyclofusion, could be due to a tendency for each tilted monocular image to normalize to the vertical.

Thus, although the tendency to see non-equidistant objects as equidistant seems to be well established as a phenomenon, it is by no means certain that a shift in the central stereoscopic disparity mechanism is involved.

This research was supported by D.S.I.R. grant No. 1208/69/1.

## REFERENCES

BERGMAN, R., and GIBSON, J. J. (1959). The negative after-effect of the perception of a surface slanted in the third dimension. *Amer. J. Psychol.*, **72**, 364–74.

DITCHBURN, R. W., and PRITCHARD, R. M. (1960). Binocular vision with two stabilized retinal images. *Quart. J. exp. Psychol.*, **12**, 26–32.

GIBSON, J. J., and RADNER, M. (1937). Adaptation, after-effects and contrast in the perception of tilted lines. I. Quantitative studies. *J. exp. Psychol.*, **20**, 453–67.

GOGEL, W. C. (1956). The tendency to see objects as equidistant and its inverse relation to lateral separation. *Psychol. Monogr.*, 70, No. 4 (Whole No. 411).

HARKER, G. S. (1962). Apparent frontoparallel plane, stereoscopic correspondence, and induced cyclotorsion of the eyes. *Percept. Mot. Skills*, **14**, 75–87.

HOWARD, I. P. (1961). An investigation of a satiation process in the reversible perspective of revolving skeletal shapes. *Quart. J. exp. Psychol.*, **13**, 19–33.

KÖHLER, W., and EMERY, D. A. (1947). Figural after-effects in the third dimension of visual space. *Amer. J. Psychol.*, **60**, 159–201.

KÜNNAPAS, T. (1960). Scales for subjective distance. *Scand. J. Psychol.*, **1**, 187–92.

LIT, A. (1959). The effect of fixation conditions on depth discrimination at scotopic and photopic illumination levels. *J. exp. Psychol.*, **58**, 476–81.

OGLE, K. N. (1946). The binocular depth contrast phenomenon. *Amer. J. Psychol.*, **59**, 111–26.

OGLE, K. N., and WEIL, M. P. (1958). Stereoscopic vision and the duration of the stimulus. *A.M.A. Arch. Ophthal.*, **59**, 4–17.

WERNER, H. (1938). Binocular depth contrast and the conditions of the binocular field. *Amer. J. Psychol.*, **51**, 489–97.

*Manuscript received 21st December, 1962.*

# SUBJECTIVE HARDNESS OF COMPLIANT MATERIALS

BY

ROLAND HARPER and S. S. STEVENS

*From the University of Leeds and Harvard University*

The apparent hardness and softness of nine samples of compliant materials were scaled by direct magnitude estimation and by cross-modal matches to the apparent force exerted on a hand dynamometer and a finger dynamometer, and to the loudness of a band of white noise. The physical hardness (force/indentation) of the compliant specimens covered a range of more than 100 to 1, extending from a fairly soft sponge to a fairly hard block of rubber. The apparent hardness of the specimens was found to follow the psychophysical power law. Subjective hardness grows as the physical hardness raised to a power. The indicated exponent was about 0·8 for magnitude estimation, about 0·7 derived by calculation from handgrip matches, and about 0·6 derived by calculation from loudness matches.

Numerical estimates and cross-modal matches for softness gave functions that were approximately the reciprocal of the functions given for hardness. Hardness is a continuum on which there exists an upper threshold.

Studies of the apparent hardness of materials that can be deformed between finger and thumb date back to 1938 when Scott Blair first undertook investigations related to the manual craftsmanship exercised by cheesemakers and cheese graders. This work was influenced by the early studies by Katz (1938) on the skills of test bakers. Harper (1952) has summarized the main developments up to 1950, including his own contributions. Most of these studies were concerned with differential sensitivity, with the ranking of stimuli, or with rating stimuli on category scales.

The present experiments applied the techniques of direct magnitude scaling and cross-modal matching to the hardness-softness continuum. Both the apparent hardness and the apparent softness of selected specimens of compliant materials were scaled by the method of magnitude estimation (S. S. Stevens, 1958). In addition, cross-modal comparisons were made between apparent hardness and three other sensory modes: (1) force of handgrip, (2) force exerted on a finger dynamometer, and (3) the loudness of white noise. Similar cross-modal comparisons were made with the apparent softness of the samples.

The purpose of these experiments was to determine how the composite tactile and kinaesthetic attribute, apparent hardness, varies with the physical hardness of rubber-like materials. The immediate question is whether, in accordance with the psychophysical law (S. S. Stevens, 1957), subjective hardness is a power function of physical hardness, and, if so, what is the exponent? The results leave little doubt about the first question: the psychophysical relation certainly approximates a power function, but the different methods used in the present series of experiments gave somewhat different estimates of the size of the exponent.

## MATERIALS AND PROCEDURE

The experimenter presented each subject with several specimens of materials which he could compress between finger and thumb. The subject was seated at a table with his eyes closed and one wrist resting on the top of a 10-in. vertically placed board that screened the stimuli from his view. He opened his hand until the experimenter touched his thumb with a sample of a material. Thereupon the subject squeezed the sample and made a judgement, with his eyes closed.

The nine specimens used covered a physical range of more than 100 to 1, but they were not all spaced in hardness as uniformly as one would wish. In the following description of the specimens, all of which were approximately 1 in. thick, the original identifying labels are used. The softest specimens, II and II(*a*), were light, close-textured sponge rubber. Specimen II was about 2 in. in diameter; II(*a*) was a piece of the same material, 4 in. square. The difference in size and shape made II(*a*) slightly harder than II, both subjectively and physically. Specimen IV was a rectangular piece of close-textured Neoprene, 1·75 in. wide. Specimen V was a rectangular piece of open-textured sponge rubber with rough textured "skin." Despite these variations in texture, these four specimens of sponge were selected because they were already available in the laboratory, and they served to extend in the soft direction the range provided by the five 2·25-in. discs of vulcanized rubber that were manufactured specially for the experiment. These manufactured specimens were labelled 20, 25, 30, 40 and 70, which were the "Shore Durometer Readings" provided by the manufacturer. Sample 70 felt hard to the touch but could still be noticeably deformed between finger and thumb, at least by the stronger subjects. Specimens having a Shore Durometer Reading greater than 70 were rejected as being too hard for the present studies.

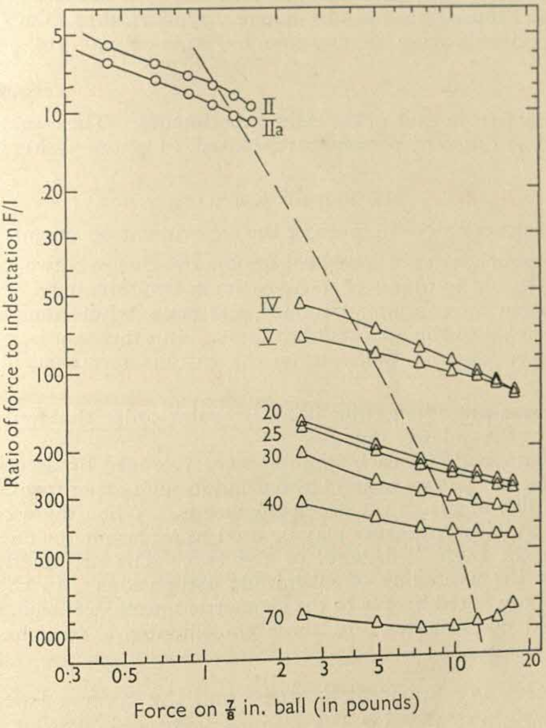


FIGURE 1

The ratio of force to indentation (F/I) used to define the physical hardness of the specimens. Since the index (F/I) does not remain constant under different applied forces, it was necessary to choose a set of arbitrary but representative points on the curves. These hardness values are given by the intersections between the dashed line and the curves. The Roman and Arabic numerals were used to designate the various samples. No. "II" was the softest specimen and "70" was the hardest.

In order to obtain comparable physical measures on all nine specimens, the ratio of force to indentation (F/I) was measured by means of a device that simulated the action of pressing a finger or thumb into a compliant mass. A similar device was developed some years ago to test the firmness of hard-pressed cheeses and has been described by Caffyn and Baron (1947). The present apparatus consisted of a counterpoised lever with a sliding weight. A steel ball, ⅞ in. in diameter, was fastened to the lever 10 in. from the pivot. This ball was pressed into the surface of the specimen by one or the other of two different weights, 0·42 lb. or 4·94 lb., set at 5, 10, 15, . . . 35 in. from the pivot. A pointer, 37·25 in. from the pivot, gave position readings on a centimetre scale. The specimen to be tested was placed on an adjustable platform immediately below the steel ball, and, with the lever in the "zero" or horizontal position, was brought into light contact with the ball.

The ratios of force to indentation (F/I) for the various positions of the load are indicated in Figure 1. Since the curves in Figure 1 are not exactly horizontal, the value of F/I does not remain precisely constant as the physical load is altered. A section (dashed line) through the family of curves was chosen to define the numerical value of the hardness of each specimen. The position of this line is essentially arbitrary, but it takes account of the fact that subjects squeeze the harder specimens with considerably more force than the softer ones. The finger dynamometer described below was used to ascertain how hard they squeezed the samples. The range of pressures was quite wide, but the median values agreed with the results shown in Figure 4. Certain other sections through the family of curves in Figure 1 were examined, but since the appearance of the plots (Figs. 2 to 5) was not greatly altered by such changes, only the ratios of force to indentation given by the dashed line in Figure 1 are considered here. Only *relative* values of F/I are required and for the nine specimens these are 8·3, 10·0, 65, 83, 230, 240, 315, 430 and 970.

### *Subjects*

Ten subjects took part in each of the eight experiments. They were selected according to availability, and 47 different persons participated, of whom 37 were males.

## Magnitude Estimation

The following instructions were given in the experiment on magnitude estimation.

> "I am going to present a series of specimens for you to squeeze between finger and thumb. Estimate the apparent hardness of each specimen by assigning a number. Use whatever numbers seem appropriate, fractions, decimals, whole numbers, but make the numbers *proportional* to the apparent hardness. In other words, the numbers should indicate *how many times* or *what fraction* the various specimens stand in relation to one another."

Each specimen was presented twice in an irregular order that was different for each subject. No explicit "standard" was used.

The 20 judgements made for each stimulus were averaged by determining the median and the geometric mean. As we nearly always find in such experiments, the medians were not systematically different from the geometric means. When these two statistics are in reasonable agreement, either measure may be used to represent the data. The geometric mean, however, is the more "efficient" of the two. The arithmetic mean is strictly contraindicated for the averaging of magnitude estimations. In the following figures straight lines have been fitted by eye to the geometric means or medians plotted in double logarithmic (log-log) co-ordinates. In these co-ordinates a straight line represents a power function.

Figure 2 indicates how apparent hardness and the inverse aspect, apparent softness, vary with the physical value F/I. Since each subject used his own arbitrary unit, or modulus, it is sometimes convenient to multiply the values by one or another constant. Thus the judgements of softness were all multiplied by 2 before they were plotted in Figure 2. This procedure improved the appearance without affecting the slope (exponent) of the function. A straight line with a slope of 0·8 has been fitted visually to the data on apparent hardness, and a similar line with reciprocal slope (opposite sign) has been fitted to the softness estimates. To the extent that the data fit these lines it means that subjective hardness grows approximately in direct proportion to the physical magnitude raised to the power 0·8, and that subjective softness grows by the same function with a negative exponent.

To a fair approximation the results in Figure 2 express a reciprocality between subjective hardness and softness: a sample that is twice as hard as another is thereby half as soft. Similar reciprocal relations have been obtained experimentally on five other continua (S. S. Stevens, 1962).

As shown in Figure 2, judgements of softness (circles) show the greater variability about the line. There is also a slight downward concavity shown by the circles in Figure 2. Both these features, greater variability and slight downward concavity, have been observed in inverse judgements on other continua. The subjects frequently report that they find it more difficult to estimate magnitudes in terms of the reciprocal function, a fact that is scarcely surprising. One would expect still more difficulty if subjects were required to report the logarithms of apparent softness or some other mathematical transformation.

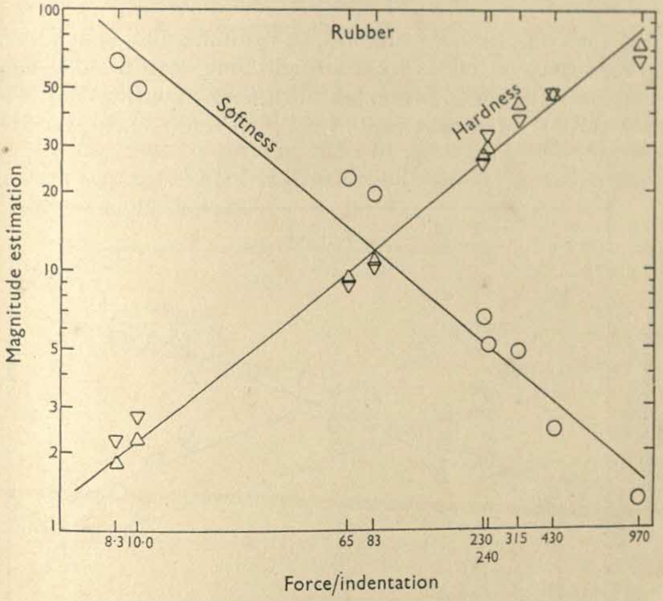


FIGURE 2

Geometric means of the magnitude estimations of subjective hardness and softness. The index of physical hardness is shown on the abscissa. The inverted triangles show the results of a repeat experiment.

Two sets of data for hardness are shown by the two kinds of triangles in Figure 2, and also in Figure 5. In order to test the reproducibility of the first experiments, the magnitude estimation of hardness and the matching of loudness to hardness were repeated 2 months later by Dr. Miguelina Guirao. The two sets of triangles demonstrate how closely the results agreed when another experimenter repeated the procedure with a group of 10 subjects, five of whom had not participated in the earlier experiment. In the repeat experiments, care was taken to balance the order: half of the subjects first judged hardness and later matched loudness to hardness; the other half did the reverse. The matching procedure is discussed below.

## CROSS-MODAL STUDIES

Three series of cross-modal comparisons were undertaken in which the nine specimens provided one of the sets of stimuli. In the first experiments, the apparent hardness (or softness) of the nine specimens was matched with the force exerted by the subject on the hand dynamometer that was developed by J. C. Stevens and Mack (1959). In the second, the apparent hardness (or softness) was matched with the force exerted on a specially designed finger dynamometer. In the third, apparent hardness (or softness) was matched with the loudness of a band of white noise. The main purpose of these studies was to test the general validity of the power law by appeal to the following argument. Consider two physical continua $\phi_1$ and $\phi_2$. With the constants that specify the size of the units omitted, the power law states that

$$\psi_1 = \phi_1^{\beta_1} \quad \psi_2 = \phi_2^{\beta_2}$$

In cross-modal matching the subject equates $\psi_1$ and $\psi_2$, a condition that is satisfied when the corresponding physical values have the relation

$$\phi_1 = \phi_2^{\beta_2/\beta_1}$$

This means that the two physical values ($\phi_1$ as ordinate and $\phi_2$ as abscissa) may be represented in log-log co-ordinates by a straight line with a slope equal to $\beta_2/\beta_1$. Cross-modal matches relating to a number of different continua have already been found to satisfy this set of requirements (J. C. Stevens, Mack and Stevens, 1960; S. S. Stevens, 1959).

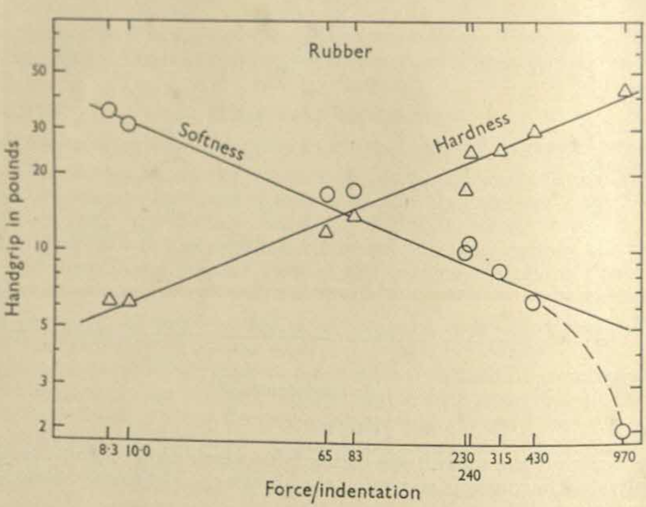


FIGURE 3

Median forces exerted on a hand dynamometer to match the apparent hardness and softness of the specimens.

*Force of handgrip matched to apparent hardness (or softness)*

The hand dynamometer was set up to the right of the subject, who squeezed its handle with his right hand while squeezing one or another of the nine specimens in his left hand. A "universal joint" connected the handle of the dynamometer to a Dillon tensile gauge graduated in 1-lb. steps from 0 to 100 lb. The 10 subjects who took part were instructed as follows:

> "I want you to match the apparent hardness of a number of specimens presented to you with the *force* you exert on the dynamometer. Compress each specimen between finger and thumb, and then match the *apparent hardness* with the force you exert on the dynamometer, and tell me when this match has been achieved. Relax your grip after each trial before the next specimen is received."

Similar instructions with the word softness substituted for hardness were used in the second part of this particular study. The specimens were presented three times to each subject, in a haphazard order, different for each series and for each subject. The first stimulus was always drawn from near the middle of the stimulus range.

The relation between the median force of handgrip and the physical hardness of the various specimens, according to the matching judgements, is shown in the log-log co-ordinates of Figure 3.

The line relating the dynamometer force to the values of F/I has a slope of 0·42 and gives a close fit to the data. A line with a slope of −0·42 has been drawn through the points corresponding to the dynamometer forces that were matched to the softness of the specimens. It is evident that this line provides a less satisfactory fit to the data than is found for the matches made between force and hardness. The lowest point for softness falls out of line with the other points, largely because some subjects matched the softness of the hardest stimuli with a squeeze too faint to register on the dynamometer. Some complained that the hardest specimen had no "softness" whatever. As in other modalities (J. C. Stevens and S. S. Stevens, 1960) account may need to be taken here of a threshold effect. This is normally done by means of an additive constant in the formula for the power law. Such a threshold effect may exist at both ends of the hardness continuum, but with these particular specimens it is more likely to have shown up at the hard than at the soft end of the continuum. This point is discussed more explicitly below.

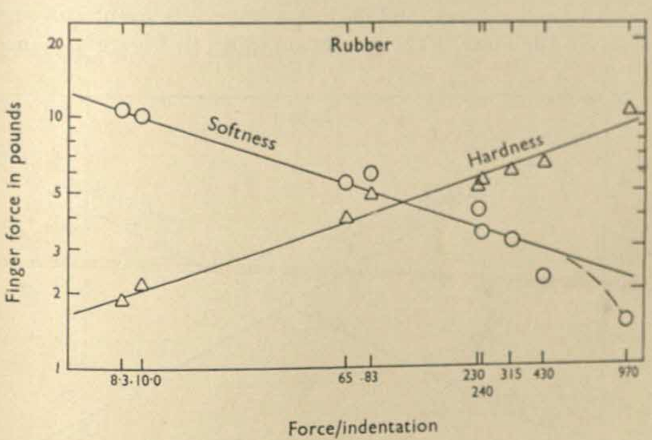


FIGURE 4

Median forces exerted on a finger dynamometer to match the apparent hardness and softness of the specimens.

*Finger force matched to apparent hardness (or softness)*

Here the method of exerting the force was modified so as to correspond more closely with the actual method of handling the compliant materials. The Dillon tensile gauge was modified by attaching two steel extensions, separated by 1 in., which could be squeezed together between finger and thumb. Since a different pivotal point was now involved, the gauge system was recalibrated by means of the same arm and weights used to measure the hardness of the nine rubber specimens. Since the response of the finger dynamometer was found to be linear, the scale readings could be converted into pounds by the multiplier 0·47. As in the preceding study, the subject used the finger dynamometer with his right hand. The instructions to the 10 subjects were unchanged from the preceding experiment, apart from requiring that each subject compress the two arms of the dynamometer between finger and thumb.

The relation between the median values of the matching force and the physical hardness of the specimens is indicated in the log-log co-ordinates of Figure 4. The slope of the line fitted by eye to the matches with hardness is 0·33, and a reciprocal slope (−0·33) has been used to represent the matches with softness. The greater scatter and the downward curvature are again evident when apparent force is matched to apparent softness.

### *Loudness matched to apparent hardness (or softness)*

With his right hand the subject manipulated a "sone potentiometer" by means of which he could vary the loudness of a band of noise (500 to 5,000 cycles per sec.) delivered through earphones. The circuit was so arranged that the experimenter could adjust the range of available intensities according to the hardness or softness of the specimen concerned. This gross adjustment was effected by means of an attenuator with 10-db. steps, which could also be used to step up the signal, when necessary, in order to measure it. The noise voltage was measured in decibels re 1 millivolt by means of a high-impedence voltmeter. The voltmeter was placed across the dynamic earphones (PDR-3 with sponge Neoprene cushions MX-41/AR). After demonstrating the controls, the experimenter gave the following instructions:

> "I want you to match the apparent hardness (or softness) of a number of specimens presented to you with the loudness of a noise. Compress each sample between finger and thumb, with the eyes closed, and adjust the intensity of the noise until a match is made. Tell me when the match has been achieved."

The nine specimens were again presented to the subject's left hand in a haphazard order, different for each subject, and each specimen was eventually matched twice with the loudness of the noise. The results are given in Figure 5. Since the decibel scale is already logarithmic, the arithmetic means of the decibel values are plotted against the logarithm of the physical hardness values (F/I) for each specimen. And since a logarithmic unit of sound pressure is 20 db. (because energy = pressure

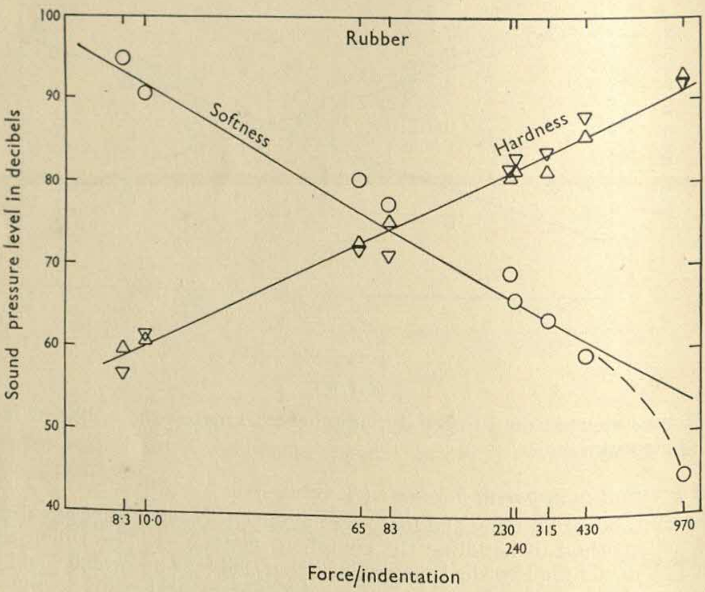


FIGURE 5

Decibel averages of the sound pressure that appeared to produce a match between loudness and the hardness and softness of the specimens. The sound pressure levels are relative to the reference 0·0002 dyne/cm². The inverted triangles show the results of a repeat experiment.

squared), the slope of the straight line fitted to the hardness matches is about 0·8. The slope of the line fitted to the softness matches is somewhat steeper, almost 1·0, but there is considerable variability about the fitted line when matches are made to softness. Furthermore Figure 5, like Figures 3 and 4, shows evidence of threshold effect for softness—indicated by a sharp downturn of the dashed curve at the low end of the softness continuum.

## Discussion

The foregoing experiments give a positive answer to the question of the applicability of the psychophysical power law to the continuum of tactile and kinaesthetic hardness. This continuum can be added to the list of more than a score of continua on which subjective magnitude has been shown to grow as a power function of the stimulus magnitude (S. S. Stevens, 1961). The logarithmic function advocated by Fechner appears to hold on none of these continua, at least when sensory magnitudes are estimated numerically or matched directly to other magnitudes.

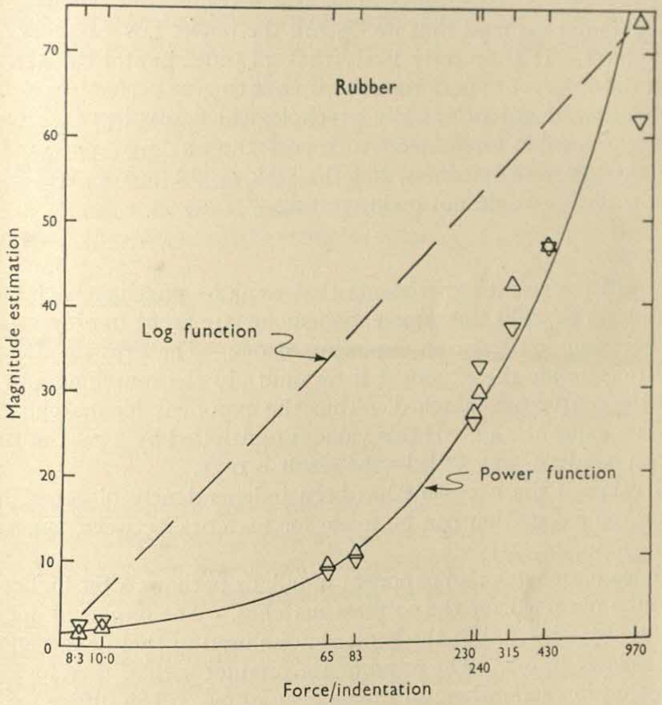


FIGURE 6

The data from Figure 2 are plotted against a linear scale of magnitude estimation and a logarithmic scale of the stimulus (abscissa). The straight line in the log-log co-ordinates of Figure 2 becomes highly curved in these semi-logarithmic co-ordinates. A logarithmic function appears as a straight line in these co-ordinates (dashed line).

In order to show how far the magnitude estimations of apparent hardness depart from a logarithmic function, the triangles that represent these estimates in Figure 2 have been replotted in semi-logarithmic coordinates in Figure 6. The curve in Figure 6 is identically the same function as the straight line in Figure 2. According to

Fechner's law, the data in Figure 6 should lie on the straight dashed line, because in semi-logarithmic co-ordinates a logarithmic function appears as a straight line. Clearly, when subjects estimate apparent hardness the results diverge widely from the function Fechner advocated.

### *Reciprocality*

The ability of subjects to judge the inverse or reciprocal aspect, softness instead of hardness, turns out to be fairly good but not perfect. As found in other studies, there is a tendency for judgements of the inverse aspect to give data that are slightly concave downward in log-log co-ordinates. This tendency appears to have been enhanced in the present experiments by the judgements of some subjects that the hardest stimulus used had none of what they would call softness. As already noted, there seems to be a threshold effect involved. Despite these problems, plus the greater inherent difficulty of the judgement, people give reasonably reciprocal functions when they are asked to judge or match in terms of the inverse of a given continuum.

The tendency of the judgements to show a second-order departure from strict reciprocality does not demand that we discard the power law governing the primary attribute, hardness. It is scarcely likely that all judgemental transformations that one could ask the observer to perform will turn out to give perfectly consistent results. Logical equivalence is not necessarily psychological equivalence. For example, it may be logically possible for subjects to report the angle whose cotangent is proportional to the apparent hardness, but the task might prove so difficult for most people that the results would not be interesting.

### *Divergent exponents*

A feature of the present experiments that remains puzzling is the failure of the various procedures to yield the same exponent for apparent hardness. The method of magnitude estimation gave an exponent of 0·8. The cross-modal matches can also be used to estimate the exponent if we multiply the matching exponent by the exponent of the continuum matched. Thus the exponent for matching to handgrip (Figure 3) has a value of 0·42. If this value is multiplied by 1·7—the value that was obtained when handgrip was scaled—the result is 0·71.

Since no value of the exponent has been independently obtained for the finger dynamometer, no calculation can be made for matches between finger squeeze and hardness.

The loudness match gave an exponent of 0·8, or perhaps a bit higher if we weight it by the results obtained for the softness matches. The exponent for the loudness of a 1,000-cycle tone is 0·6, but the band of noise used in these experiments typically gives a value closer to 0·7. The appropriate product yields, therefore, an exponent for hardness that lies somewhere in the vicinity of 0·6. The difference between this estimate and that produced by direct magnitude estimation, 0·8, is greater than any discrepancy yet observed. (For a list of nine such comparisons, see S. S. Stevens, 1961.)

Why the large discrepancy when loudness is matched to hardness? We do not know for certain, but there is this to be said. The squeezing of a rubber-like substance gives sensations that have many aspects—touch, pressure, muscular exertion, and so forth. As with the silicone resin known as "bouncing putty," different properties stand out, depending on what one does. It is possible that matching one kind of stimulus to apparent hardness calls attention to one aspect of the situation, and that matching another kind calls attention to some other aspect. It is even possible that a person squeezes the compliant specimen differently, depending on whether he is trying to match its hardness to a number, to another pull, or to a loudness. In any case, against the background of more than a score of crossmodal matches that have shown excellent agreement with prediction, it is interesting to have found a kind of stimulus for which the exponent derived from cross-modal matches does not agree with that predicted from direct magnitude estimations. It poses a puzzle.

In all the experiments reported above, the values determined for the exponent of hardness are strikingly lower than the value 1·7 obtained by J. C. Stevens and Mack (1959) for subjective force of handgrip. Although both operations involve squeezing, the perception of hardness seems to have a different basis from the perception of muscular effort.

### *Thresholds*

As noted above, the hardness-softness continuum exhibits another novel feature, namely, evidence of an upper threshold. There is also a lower threshold, presumably, but the hardness of our samples did not extend down to its vicinity. When an experiment needs to take explicit account of the lower threshold, $\phi_0$, the power law relating the psychological magnitude $\psi$ to the physical magnitude $\phi$ can be written

$$\psi = \mathrm{k}(\phi - \phi_0)^\beta$$

This equation says in effect that, in order for the power law to hold, the stimulus must be measured starting at the threshold value. The continuum for the apparent warmth of objects applied to the forearm provides an obvious example of the need to measure the stimulus as the difference between the threshold or "neutral" temperature and the temperature of the stimulus object (J. C. Stevens and S. S. Stevens, 1960).

When, as with hardness, there exist stimulus values above which there are no corresponding gradations in the subjective magnitude, the problem of an upper threshold, $\phi_u$, enters the picture. It is conceivable that the power law may then take the form

$$\psi = \mathrm{k}\left(\frac{\phi - \phi_0}{\phi_u - \phi}\right)^\beta$$

A graphical representation of this equation, with an arbitrary set of parameters, is shown in Figure 7. The lower threshold $\phi_0$ was set at 1 and the upper threshold $\phi_u$ at 1,000. The exponent was given the value 0·5. When a power function has this form, the curve is limited asymptotically at the two ends where it approaches the values of $\phi_0$ and $\phi_u$.

When this model is applicable, the curve for the inverse function, e.g. softness instead of hardness, resembles Figure 7 turned upside down. In the foregoing experiments, although care was taken to eliminate the rubber specimens that felt "infinitely" hard, i.e., those that had no detectable softness, it appears that for some subjects specimen "70" was near or above the upper threshold. The tendency for this result to show up most clearly in the judgements of softness suggests that the nature and location of the upper threshold, when it exists, may perhaps best be studied in terms of the inverse aspect. The dashed portions of Figure 3, 4, and 5 were drawn to illustrate this point.

The inverse effect, an upward curvature at the right-hand end of the curve for hardness, did not manifest itself in the experimental data. There are several possible reasons for this fact, among them the difference in the subjects' approach to each of the two tasks, his normal resistance to using very large numbers, and his aversion to producing extremely intense stimuli in cross-modal matching. In other words, it may be that the symmetry depicted in Figure 6 cannot be easily duplicated in an experiment, even if the range of stimuli could be made to extend from the lower to the upper threshold.

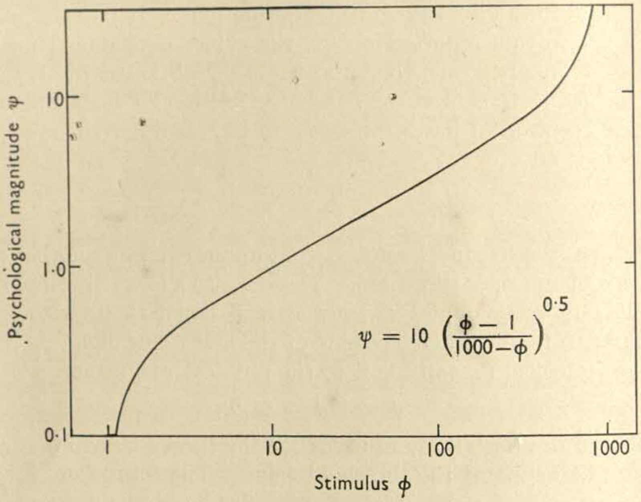


FIGURE 7

An example showing how the power function may appear in log-log co-ordinates when there is an upper as well as a lower threshold.

It should be noted that no upper threshold has thus far been found for any of the simpler sensations, loudness, brightness, saltiness, etc. On the other hand, it may prove possible to demonstrate an upper as well as a lower threshold on certain other continua. Viscosity and inertia are two continua that suggest themselves.

*The continuum*

Two of the compliant specimens used as stimuli happened to be very close together in hardness: the values of the ratio F/I were 230 and 240. In the total of ten experiments depicted in the fore-going figures, the geometric means show that the physically harder specimen was judged harder (or not softer) in all but one instance, namely, when handgrip was matched to apparent softness (Fig. 3). This matter is mentioned here in order to emphasize the nature of these sensory continua. To the extent that the continua are in fact continuous, there is in principle no limit on how close two stimuli may be spaced and still be resolved. The averaging of sufficient judgements will detect the difference between them and place them correctly on the scale. In this sense there is no just noticeable difference, for there is no specifiable separation at which two stimuli would fail to be properly resolved if geometric means were taken on a large enough sample of data. The conventional just noticeable difference is merely an arbitrary measure of the scatter or dispersion among people's discriminatory judgements.

This research was supported by Grant B-2974 from the National Institutes of Health and Grant G-10716 from the National Science Foundation (Psychophysics Laboratory Report PAR-273). The study was carried out while Roland Harper was visiting Harvard University during a period of Study Leave from the University of Leeds.

## REFERENCES

American Society for Testing Materials. Standard methods of test for hardness of rubber. ASTM Designation: D 314–58, 1958, 1297–1298.

CAFFYN, J. E., and BARON, M. (1947). Scientific control in cheesemaking. *The Dairyman*, **64**, 345–6.

HARPER, R. (1952). Psychological and psychophysical aspects of studies of craftsmanship in dairying. *Brit. J. Psychol. Monogr. Suppl.*, **28**, 1–63.

KATZ, D. (1938). The judgements of test bakers. A psychological study. *Occup. Psychol.*, **12**, 139–48.

STEVENS, J. C., and MACK, J. D. (1959). Scales of apparent force. *J. exp. Psychol.*, **58**, 405–13.

STEVENS, J. C., MACK, J. D., and STEVENS, S. S. (1960). Growth of sensation on seven continua as measured by force of handgrip. *J. exp. Psychol.*, **59**, 60–7.

STEVENS, J. C., and STEVENS, S. S. (1960). Warmth and cold: Dynamics of sensory intensity. *J. exp. Psychol.*, **60**, 183–92.

STEVENS, S. S. (1957). On the psychophysical law. *Psychol. Rev.*, **64**, 153–81.

STEVENS, S. S. (1958). Problems and methods of psychophysics. *Psychol. Bull.*, **54**, 177–96.

STEVENS, S. S. (1959). Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.*, **57**, 201–9.

STEVENS, S. S. (1961). Psychophysics of sensory function. In W. A. ROSENBLITH (Ed.), *Sensory Communication*. New York.

STEVENS, S. S. (1962). Sensation measurement by inverse judgments. *Science*, **136**, 329 [Abstract].

*Manuscript received 1st January, 1963.*

# SUBJECTIVE EXPECTANCY AND CHOICE REACTION TIMES

BY

G. H. MOWBRAY

*From the Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland*

Previous findings suggested that selective response times might be affected both by the inter-stimulus interval and by the probability of occurrence of the stimulus for reaction. These two factors have been tested independently and have been found to influence reaction times in a fashion that an expectancy hypothesis would predict.

In a recent series of experiments concerning the effect of number of alternatives on choice reaction time for overlearned stimulus-response combinations, it was shown that the two classical types of choice reaction produce dissimilar results under identical stimulus conditions (Mowbray, 1960). In summary, the case that calls for a different response to each of a number of different stimuli, provided reaction times that did not increase as a function of the number of alternatives. However, the case that calls for a response to only one particular stimulus—all others being ignored—provided reaction times that increased linearly as the number of alternative stimuli was increased. Classically these conditions are known as the Donders b-reaction and the Donders c-reaction, respectively. In contemporary usage the former is termed choice or disjunctive reaction, and the latter, selective response (Brebner and Gordon, 1962).

Physically, the experiment using the choice reaction was identical to that using the selective response, with one major exception. In the former case, subjects were required to respond to every stimulus, while in the latter case, a response was required only when the key stimulus appeared. Therefore, in one instance, responses occurred regularly every 2·5 sec., but in the other instance responses occurred irregularly with the time interval between responses increasing on the average as the number of choices increased. In order to determine whether the increasing inter-stimulus interval had contributed to the differences observed, an analysis of the reaction times to the key stimulus as a function of this interval was made. First of all it was revealed that the distributions of inter-stimulus intervals were not the same for all conditions of choice. The conditions employing smaller numbers of choices had fairly regular, normal distributions of interstimulus intervals, i.e. there were a few short intervals and a few long intervals with the majority falling somewhere between in a more or less regular fashion. However, the conditions that employed the larger number of choices had much less symmetrical distributions. They were, in fact, highly irregular. Furthermore, the analysis showed also that for those conditions having regular and normal distributions of inter-stimulus intervals reaction times decreased as the intervals became longer. No such trend was present when the distributions of intervals were irregular.

On the strength of this analysis it was tentatively proposed that the responders were influenced by expectancies derived from what they conceived to be the structure of the stimulus situation. These expectancies in turn influenced the choice reaction times, causing them to be long when expectancies were low and to decrease gradually as expectancies became higher. No more is implied here by expectancy than the subjective estimation of the time at which a stimulus calling for a response is likely to occur. Presumably, as the subject gains experience with the task, he develops some notions about the statistical structure of it and tailors his behaviour to suit. Furthermore, it seems likely that at least two factors can influence expectancies. These are the distribution of inter-stimulus intervals and the number of stimuli that require a response relative to those that do not require a response. As to the first of these, it is proposed that when stimuli requiring a response are distributed throughout a sequence in such a way that many occur in close succession and few are separated by long intervals, subjects maintain an alertness and a consequent lowering of reaction times that is absent in the converse situation. This alertness will be reflected not only in the mean, overall reaction times but individual reaction times will follow to a certain extent the actual distribution of intervals employed. In other terms, reaction times will be lower at the modal points of the distribution of inter-stimulus intervals than at other points. The second factor, while akin in many ways to the first, is yet conceived of as exerting its own special influence. Given identical distributions of inter-stimulus intervals and equal probability of occurrence of all stimuli, a two-choice reaction time task will require a response to 50 per cent. of all stimuli, whereas a ten-choice task will require a response to only 10 per cent. of the stimuli. In the former case, a greater degree of alertness will be maintained with a consequent reduction of reaction times relative to the latter.

The following experiments were designed to test independently the effect of these two factors.

## EXPERIMENT I

The first experiment is an investigation of the effect on the selective response latency of various distributions of inter-stimulus intervals, holding average rate of stimulus occurrence constant.

### METHOD AND PROCEDURE

The apparatus employed in this research has been described in detail elsewhere (Mowbray, 1960). Briefly, it is capable of displaying any one of the 10 Arabic numerals, 0–9, at any reasonable speed, as programmed from a punched-tape, teletype transmitter. The reaction time measured was the time elapsed between the appearance of a numeral and the vocal naming of the numeral by the subject. An electronic voice key was used in conjunction with a Standard Timer, Model S-1.

For this research, all 10 numerals were used with each numeral appearing with approximately equal frequency at 2·5 sec. intervals. Three sets of tapes were prepared in such a manner that the key stimulus (the figure 8 in every case) had a different distribution of inter-stimulus intervals in each set. The individual sets contained approximately 1,300 stimuli. For ease of storage and handling, all sets were divided into 13 tapes of about 100 stimuli each.

The three distributions of intervals employed were (*a*) normal, (*b*) rectangular, and (*c*) a "J" distribution. Figure 1 shows a plot of the distributions with the number of times that a given interval was utilized. It will be noted that the number of key stimuli and the total number of stimuli for each distribution is not exactly equal This compromise was consciously adopted after many attempts to obtain reasonably adequate "J" distributions that matched the other distributions in total numbers of stimuli having equal probabilities of occurrence. An ideal solution to the problem can be made by use of Monte Carlo statistics, but it was felt that such precision was not required. The probability of occurrence of a key stimulus was thus about 0·08 for the normal and the rectangular distribution and 0·11 for the "J." It was assumed that this slight deviation from equal probability of occurrence (0·10) would in no way affect the results.

The subjects were 12 college students between the ages of 19 and 38 who had served previously in a voice reaction time experiment. They were selected from among a total of 150 subjects on the basis of the homogeneity of their variances for choice reactions. There were six males and six females. Each subject provided choice reactions for all three of the distribution sets, i.e. they responded to all stimuli as they appeared. Following this, they provided selective responses (responding only to the appearance of the figure 8) for the three sets of distributions. In this latter case, each distribution set was presented twice to each subject. The order of presentation of the distribution sets was different for each subject. One experimental session lasted approximately 1 hr. and consisted of the presentation of one distribution set. Thus each subject served in nine experimental sessions and contributed a total of about 4,700 reaction times. In the analysis only the reaction times to the key stimuli were used, and these were all reduced by 65 millisec. from the recorded time since that was the measured delay caused by relay inertia of the apparatus.

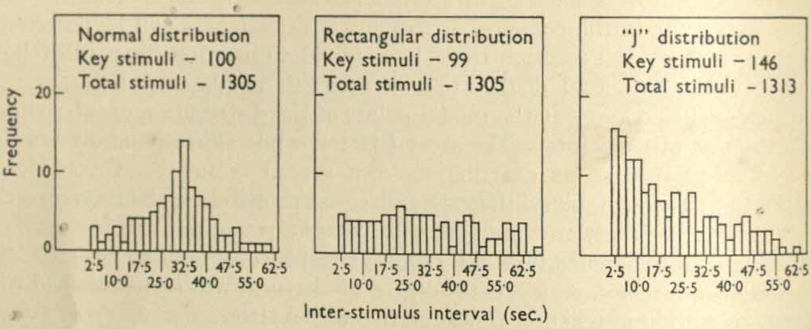


FIGURE 1

The distributions of inter-stimulus intervals employed for all three distribution sets.

## RESULTS

The first results to note are the mean reaction times for the two types of responses and for the three distributions of inter-stimulus intervals. These are shown in Table I, along with the standard deviations and the total number of reactions contributing to the mean values. It is of importance to observe that the selective responses yielded higher reaction times than the choice reactions as reported earlier for the 10-choice condition (Mowbray, 1960). Furthermore, the choice reactions did not appear to be affected by the distributions of the inter-stimulus intervals of the key stimuli. The reaction times using the "J" distribution, while lower than the other two, did not show a significant difference when tested with Student's *t* for correlated scores. On the other hand, the distributions of the key stimuli for the selective responses did significantly affect the latencies. The difference between the

TABLE I

MEAN REACTION TIME, STANDARD DEVIATIONS AND N'S FOR THE THREE DISTRIBUTIONS OF INTER-STIMULUS INTERVALS AND FOR TWO TYPES OF REACTIONS (12 SUBJECTS)

| *Reaction—type and distribution* | *Mean reaction time (sec.)* | *Standard deviation* | *Number of reactions* |
|---|---|---|---|
| Choice reaction | | | |
| Normal distribution . . . . | 0·430 | 0·069 | 1,150 |
| Rectangular distribution . . | 0·431 | 0·041 | 1,161 |
| "J" distribution . . . . | 0·425 | 0·068 | 1,716 |
| Selective response | | | |
| Normal distribution . . . . | 0·469 | 0·085 | 2,324 |
| Rectangular distribution . . | 0·478 | 0·092 | 2,314 |
| "J" distribution . . . . | 0·452 | 0·081 | 3,383 |

normal distribution and the rectangular distribution was tested for significance (again with Student's *t*) and it was found that such a difference could occur by chance less than one time in a thousand. The "J" distribution, which gave the lowest reaction times of all, was of course significantly different from the rest.

It is perhaps not unexpected that the standard deviations for the choice reactions should be consistently lower than those for the selective responses. The irregularity of the latter must provide some amount of uncertainty for the subject, leading in turn, to more variable performance.

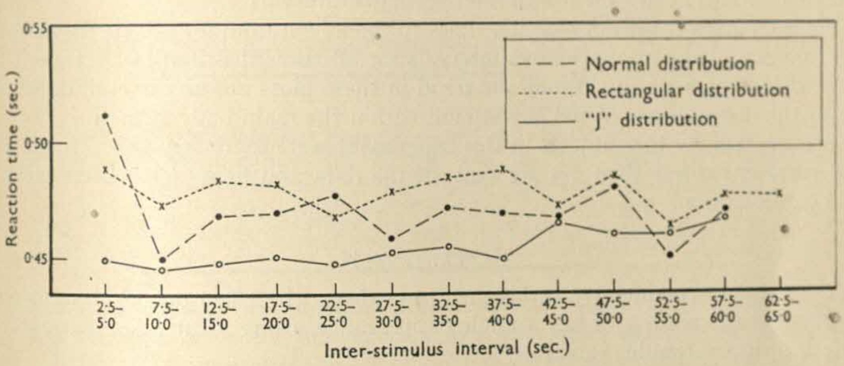


FIGURE 2

Mean selective response latencies as a function of the interval in seconds between key stimuli for the three distribution sets.

The second results to note concern the effect of the inter-stimulus interval on the individual reaction times. The purpose of this analysis is to determine whether the reaction times reflect the realities of the stimulus sequence. Figure 2 shows mean reaction times plotted as a function of the inter-stimulus intervals for all three distributions.

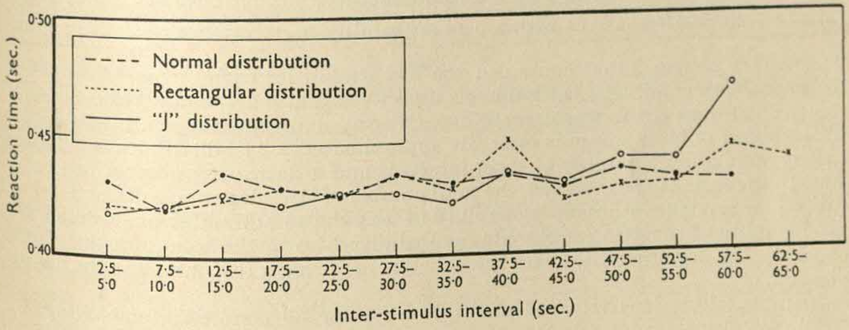


FIGURE 3

Mean choice reaction times as a function of the interval in seconds between key stimuli for the three distribution sets.

It should be pointed out that successive intervals have been combined both for ease of plotting and for the stability gained for those intervals that contain relatively few reaction-time measurements.

Looking first at the plot of the "J" distribution data, it is evident that the shortest reaction times are associated with the mode of the distribution of inter-stimulus intervals. The large number of short intervals relative to longer ones apparently served to maintain an alertness in the responder that is reflected by his reaction times.

The data for the normal distribution are not so clear-cut. While it is true that reaction times for the short intervals, which were relatively rare, are considerably raised, the shortest times are not clearly associated with the most numerous intervals. Rather, they are displaced towards the short interval end of the distribution.

Finally, for the rectangular distribution where there is no clear modal pattern of inter-stimulus intervals, reaction times generally reflect that fact. In other words, they remain fairly constant for all inter-stimulus intervals.

Figure 3 gives the mean reaction times to the key stimuli for the choice reactions as a function of the inter-stimulus intervals for all three distributions. Here it can be seen that there was no systematic trend in these plots nor any overall differences due to the distributions of the key stimuli within the main body stimuli.

Errors made by the subjects in this experiment were extremely few. In all cases they represented less than 0·5 per cent. of the data and hence have been excluded from consideration.

## EXPERIMENT II

The second experiment was designed to evaluate the effect on selective response latency of various rates of key stimulus appearance, using identically shaped distributions of inter-stimulus intervals.

### METHOD AND PROCEDURE

The apparatus and the general method employed were the same as those of the previous study; only the independent variables were different.

Two separate choice conditions were used in conjunction with two probabilities of occurrence of the key stimulus to provide the following four experimental conditions:

(1) Four-choice alternatives with a 0·10 probability of occurrence of the key stimulus.

(2) Four-choice alternatives with a 0·25 probability of occurrence of the key stimulus. In other words, all four alternatives were equally likely.

(3) Ten-choice alternatives with a 0·10 probability of occurrence of the key stimulus.

(4) Ten-choice alternatives with a 0·25 probability of occurrence of the key stimulus.

Tapes were prepared that contained 100 key stimuli for each of the four conditions. In the four-choice conditions the numerals used were 1, 4, 6 and 8, and, of course, in the 10-choice conditions the numerals were 0–9. For all experimental conditions employed, the distribution of inter-stimulus intervals approximated a "J" distribution. That is to say, there were a large number of short intervals and a decreasing number of long ones. Because of the restrictions involved, the total number of stimuli per tape differed. For the conditions employing a stimulus probability of 0·25 the total number of stimuli was 400. For the conditions involving a stimulus probability of 0·10 the total number of stimuli was 1,000. The key stimulus, hence the only one requiring a response, was the figure "8" in all cases.

Experimental sessions required about 20 min. for all conditions employing a 0·25 probability and just under 1 hr. for the conditions employing a 0·10 probability. Once an experimental session had begun there were no interruptions until the complete series had been run.

A total of 114 subjects participated in this experiment with each subject being assigned randomly to one of the four experimental conditions. The only restriction on randomness of assignment was the requirement that an equal number of males and females be represented in each of the four conditions. Due to scheduling errors there were 15 male subjects in the four-choice condition where probability was 0·25, and 15 females in the 10-choice condition where probability was 0·25. Otherwise, each condition was represented by 14 male and 14 female subjects. With the exception of one high school senior the subjects were all college students. The age range was 17–33 years with only one subject over 28.

The subjects were not informed about the purpose of the experiment nor did they know beforehand how long the experimental session would last.

## RESULTS

The overall mean reaction times for the various experimental conditions are shown in Table II. All 100 reactions per subject, excluding any error, were used in the analysis. There are a total of 11,222 reaction times represented in this table, or approximately 98·5 reaction times per subject out of a possible 100.

TABLE II

MEAN CHOICE REACTION TIME IN SECONDS FOR DIFFERENT NUMBERS OF ALTERNATIVES AND FOR DIFFERENT PROBABILITIES OF OCCURRENCE OF THE KEY STIMULUS (114 SUBJECTS)

| *Probabilities of occurrence* | *Number of alternatives* | |
|---|---|---|
| | 4 | 10 |
| 0·10 | 0·523 | 0·534 |
| 0·25 | 0·503 | 0·491 |

The first thing to notice, and the most important result of this research, is the fact that reaction times for four choices and 10 choices are not different when the probability of occurrence of the key stimulus is the same. In other words, when the average rate of occurrence of the stimulus requiring a response is held constant, the number of different stimuli *not* requiring a response with which it is associated, has no effect on reaction time. The small differences between the means for the four-choice 0·10 probability condition and the 10-choice 0·10 probability condition were tested for significance using Student's *t*-test and were found to be not significantly different from chance. The same test was applied to the difference between the means of the four-choice 0·25 probability and the 10-choice 0·25 probability conditions with the same result.

Comparisons were also made between the 10-choice 0·10 probability and the 10-choice 0·25 probability conditions, and these means were found to be different with a "*p*" greater than 0·01. However, the same comparison for the four-choice conditions did not yield a significant difference.

Finally, diagonal comparisons between the means all yield differences that are significant in the direction required for the suggested hypothesis.

## DISCUSSION

In measuring choice reaction times of the type known classically as the Donders c-reaction, and more recently as selective response times, there are three main things to be considered.

(1) The overall, mean reaction times are influenced by the nature of the distribution of inter-stimulus intervals. Reaction times are highest for rectangular distributions, i.e. when long and short intervals between key stimuli are equally represented in the experimental sequence. Reaction times are lowest for "J" distributions, i.e. when there are a great many short intervals and relatively few long ones in the experimental sequence.

(2) Not only are over-all, mean reaction times influenced by the shape of the inter-stimulus interval distribution, but so also are individual reaction times, in that they tend to reflect the actual, physical characteristics of the distribution. Thus

it is that the lowest reaction times for the "J" distribution occur at the shortest intervals, which also are the most numerous. For the rectangular distribution where all intervals are represented an approximately equal number of times, no clear and consistent trend is in evidence. It would indeed be gratifying if reaction times for the normal distribution of intervals had adhered as closely to predictions stemming from an expectancy hypothesis as the other distributions used in this research. In one respect, it clearly can be claimed that they did. The longest reaction times are unequivocally associated with the shortest inter-stimulus intervals which in turn are the least numerous. The disappointment follows from the fact that the shortest reaction times are not associated with the most numerous inter-stimulus intervals but are, instead, displaced toward the short interval end of the distribution.

(3) Finally, the frequency with which a stimulus calling for a response occurs, relative to those not calling for a response (or the response/no response ratio) has a great deal to do with determining the size of the reaction time. When this ratio is the same, even though the number of alternatives involved varies greatly, the reaction times will be the same. This fact goes a long way toward explaining an anomalous result found in an earlier research (Mowbray, 1960). Here, it will be remembered, selective response times increased linearly as a function of the number of alternatives, and further, reaction times for large numbers of alternatives were higher than for the same number of alternatives utilizing the choice reaction. In the earlier research, not only were the distributions of inter-stimulus intervals contributing to the inflated reaction times, but so, too, were the response/no response ratios. Classically, it has always been maintained that selective response times are shorter than choice reaction times on the grounds that producing one response from a repertoire of responses requires more time than matching a single response to one of a repertoire of stimuli; in short, that it is more time-consuming to select and produce a response than it is to select a stimulus for response (Woodworth, 1938, p. 303). Faced with the evidence gathered here, one begins to wonder about the justification for such a position. In Mowbray's (1960) earlier experiment, selective response times were lower than choice reaction times for two and four equally likely alternatives, they were the same for six alternatives, and they were higher for eight and 10 alternatives. Again in Experiment I of this research, 10 equally probable alternatives yielded longer latencies with the selective response than with the choice response. However, as Experiment II has shown, the difference originally found between four and 10 alternatives using a selective response, disappears when the response/no response ratios are equated. It thus becomes evident that more is involved in the comparison than a simple statement of the type of reaction called for.

Brebner and Gordon (1964) report confirmation of the effect of stimulus probability on selective response latencies. Using a manual response to the appearance of Arabic numerals, response latencies were the same for two alternatives and four alternatives as long as stimulus probabilities were equal. On the other hand, with equal numbers of alternatives the shortest latencies were associated with the highest probabilities of stimulus occurrence.

These data confirm the suggestion made earlier that selective response times are influenced by subjective expectancies derived from the experimental task (Mowbray, 1960). These expectancies can be influenced by the way the key stimuli are distributed within a series or by the ratio of key to non-key stimuli. In the former, subjects infer, presumably on the basis of experience with the task, something about the form of the distribution of key stimuli and tend to be more alert during those portions of an experimental sequence when a response is likely to be called for. That their inferences are not always entirely accurate is perhaps not surprising.

Apparently, their estimates for the normal distribution erred on the side of brevity. Whether, with longer sequences, they would have adjusted their expectations to coincide with the facts of the situation cannot of course be ascertained. In any case, it is generally agreed that the duration of long temporal intervals tends to be over-estimated when the subject is engaged in a boring and monotonous task (Woodrow, 1951).

Deese (1955) earlier proposed an expectancy hypothesis to account for some of the results from experiments on watch-keeping behaviour. The similarity between tasks of that sort and the ones discussed above is marked. Both are concerned with the detection of an unpredictably occurring signal amidst "noise." Whereas vigilance experiments often count time between signals in terms of many minutes or possibly even hours, it is seldom more than seconds for selective response experiments. It is interesting to consider that the same principles may apply in both cases. Of further interest is the recent finding that time-uncertainty with regard to the appearance of a choice reaction signal, under certain conditions, affects the choice reaction time in a fashion that is consistent with the expectancy hypothesis herein adopted (Bertelson and Boons, 1960, 1961). They reported that constant fore-periods of 0·5 and 2·5 sec. resulted in lower choice reaction times than fore-periods varying randomly between 0·25 and 5·5 sec. in the former case, and between 0·5 and 2·5 sec. in the latter case. However, constant fore-periods of 5·0 sec., for some unaccountable reason, produced the longest choice reaction times of all. The significance of this result obviously merits further study.

This report was prepared under Contract NOw 62-0604-c between the Bureau of Naval Weapons, Department of the Navy, and The Johns Hopkins University.

## REFERENCES

BERTELSON, P., and BOONS, J. P. (1960). Time uncertainty and choice reaction time. *Nature*, **187**, 531–2.

BERTELSON, P., and BOONS, J. P. (1961). L'Influence de l'incertitude temporelle sur le temps de réaction de choix. *L'Année Psychol.*, **61**, 361–76.

BREBNER, J., and GORDON, I. (1962). Ensemble size and selective response times with a constant signal rate. *Quart. J. exp. Psychol.*, **14**, 113–6.

BREBNER, J., and GORDON, I. (1964). The influence of signal probability and the number of non-signal categories on selective response times. *Quart. J. exp. Psychol.* **16**, 56–60.

DEESE, J. (1955). Some problems in the theory of vigilance. *Psychol. Rev.*, **62**, 359–68.

MOWBRAY, G. H. (1960). Choice reaction times for skilled responses. *Quart. J. exp. Psychol.*, **12**, 193–202.

WOODROW, H. (1951). Time perception. In S. S. STEVENS (Ed.) *Handbook of Experimental Psychology*, New York.

WOODWORTH, R. S. (1938). *Experimental Psychology*. New York.

*Manuscript received 11th February*, 1963.

# THE EFFECT OF REPETITION OF ITEMS ON TWO PHASES OF PAIRED ASSOCIATE LEARNING

BY

PETER B. WARR

*From the Department of Psychology, University of Sheffield*

An investigation is reported in which the effect of repetition of items on two learning tasks is determined. These tasks are considered to be analogous to the response-learning and associative phases of paired associate learning. The findings tend to support the hypotheses under investigation, that response learning is an all-or-none occurrence and that the associative phase may be explained by the dual factor hypothesis.

## INTRODUCTION

Underwood, Runquist and Schulz (1959), Underwood and Schulz (1960), Postman (1961), Horowitz (1961) and others have argued that the standard paired associate learning task involves two phases. The learning of a paired associate item is taken to involve a response-learning phase and an associative phase. Similar conceptions involving a greater number of stages have been suggested (Cook and Kendler, 1956; Murdock, 1959; McQuire, 1961), but these agree in acknowledging the existence of response-learning and associative phases. The simpler analysis of paired associate learning into two phases will be considered here, without in any way attempting to judge the merits of the various interpretations. It is felt that the conclusions reached below are equally applicable to the multi-stage analyses of paired associate learning insofar as these accept the existence of a response-learning and an associative phase.

The term response-*learning* is used here in a wide sense: not only has a previously unencountered response term to be learned (a matrix of dots, for instance), but a familiar adjective occurring as a response term has to be discriminated from other adjectives. This is to say that response-learning still takes place when the responses are familiar terms, since the subject has to learn which of these familiar terms are relevant to his task. The associative phase is taken to involve the connecting of response terms, which have been accepted as part of the current population of such terms, to the appropriate stimulus terms. It must not be assumed that the associative phase always occurs on a trial after response-learning, although this may sometimes be the case.

The experiments to be reported here deal with the effect of repeated presentation of items on these two phases of paired associate learning. It has been argued (e.g. Rock, 1957; Rock and Heimer, 1959; Clark, Lansford and Dallenbach, 1960; Estes, 1960) that repetition is of importance in learning only in so far as it strengthens associations which have already been formed on a single trial. Learning is considered to take place on a single trial so that repeated presentation of items cannot lead to a gradual building up of associations. Warr (1963) has claimed that the results of these experiments do not provide satisfactory evidence for this argument, and has suggested that paired associate learning in some conditions requires a dual-factor hypothesis. This involves the notions that all-or-none and incremental factors may both contribute to the learning of a paired associate list and that subjects tend to "select" certain items from that list. In some circumstances, however, learning may be possible after only a single "selection," in which cases incremental learning is not necessary.

The data which provide evidence for the dual-factor hypothesis were also analysed (Warr, 1963) in such a way as to generate hypotheses about the response-learning and associative phases of paired associate learning. It was postulated that the response-learning phase may be accounted for by the all-or-none hypothesis and that the associative phase may be interpreted on a dual-factor basis. Since the evidence advanced in favour of these suggestions at the time was somewhat circumstantial, the present experiments were carried out to see if more compelling evidence for the hypotheses might be forthcoming.

The experimental design is basically that of the previous experiment. Items are removed from the list on the criterion of a single correct recall by the subject, and new items are inserted in the list in place of the learned ones. In this way the length of the list remains constant and all items presented are as yet unlearned. Such a procedure allows the probability of correct recall to be related directly to the number of presentations that an item has received.

## EXPERIMENT 1

The purpose of this experiment was to study the effect of repetition of items on response-learning. To this end, subjects were required to learn a list of two-digit numbers (which comprised the response terms in the original experiment). Learning was assessed by a free recall test after each presentation of a list. This task is considered essentially identical to the response-learning phase of paired associate learning, since the question of associating the items to a stimulus term does not arise. The subject's learning task is to discriminate the items in any one list from the remaining items of a population which is known.

### *Apparatus and Method*

Two In-Line Digital Display Units were mounted together to expose two-digit numbers in the centre of a 2 ft. by 1 ft. black screen. The subject sat facing the screen in a fairly dark room. The numbers were exposed by means of switches operated from behind the subject, at a rate of one presentation every $2\frac{1}{2}$ sec. The number was exposed for approximately $1\frac{1}{2}$ sec. after which the screen was blank for approximately 1 sec. whilst the switches were moved to the next position. During presentations of a list the subject repeated each item as it appeared, and at the end of the presentation he recalled as many items (in any order) as he could. No knowledge of results was provided.

A pilot study revealed that subjects tended to recall immediately the last three or four items of the list in the order in which these had been presented. They then tended to pause, after which they recalled isolated terms from the rest of the list in an apparently random order. It was felt that they were first emitting responses held in a short-term store before emitting those terms which could be considered examples of more long-term retention. This distinction between the phenomena of retention and the "memory after-image" (*Erinnerungsnachbild*) has long been recognized (e.g. Fechner, 1860; Titchener 1915). Whilst there is no evidence that a measure of immediate memory would yield different results to a measure of more long-term retention, it was considered preferable to attempt to keep the two measures separate.

It is possible to avoid measuring immediate memory by increasing the presentation-recall interval. But there is no guarantee that the usual ways of occupying the subject during this interval are preventing selective rehearsal of certain items from the list. Such selective rehearsal would tend to confuse any established relationship between number of presentations of an item and the probability of recall of that item.

In addition to immediate memory and selective rehearsal, a third possible artefact was suggested by the pilot experiment. In a task of this kind, subjects tend to recall correctly the last few items of a list, but are less successful on earlier items. Consequently, if it were to happen that during the course of this experiment a disproportionate number of items being presented for the first time (for instance) appeared as the last item in the list, this would artifically increase the established probability of recall after one presentation. To ensure that such a factor is not influencing the results, it is necessary that an equal proportion of the items receiving each number of presentations occurs in each serial-order position. For example, if 640 items are presented once throughout an experiment employing an eight-item list, the number of these items in each serial-order position must not depart significantly from 80. And a similar control must be exerted over items receiving all other numbers of presentations. A determined attempt was made to balance this proportion of items in each serial-order position, but since no guarantee was available in advance of the experiment that this control would be complete, it was considered advisable to employ a technique which would minimize the effect of unbalanced proportions and the possibility of the artefact discussed above.

In order to overcome these three difficulties, an experimental procedure was devised which involved presenting a 12-item list, of which only the first eight items were scored. Subjects were of course not aware that this procedure was being followed. The last four items of a list were therefore "ballast" items. Each ballast item was presented only once. The experiment comprised eight presentations and eight recall tests, so that 32 ballast items were presented to each subject. Twenty-five of these were numbers haphazardly selected from the range 12–98, excepting double numbers (22, 33, etc.) and numbers including a zero (20, 30, etc.). The remaining seven ballast items were numbers including a zero. The position of each of the ballast items was held constant for all subjects.

The first eight items in each list presented were taken from the residue of the range 12–98, with the exception of double numbers and numbers including a zero. The experimental procedure has been described elsewhere (Warr, 1963). All items correctly recalled from the experimental list (i.e. excluding the ballast items) were replaced by new items haphazardly selected from the same population. In this way the experimental list was determined by the subject's learning of the previous list and by the selection of new items. Order of experimental items was varied between presentations. On the other hand, the position of the four ballast items in each presentation was held constant for all subjects. It must be emphasized that no subject was aware that this procedure was being followed. Although it is possible that the ballast items may interfere with other items in any list this interference will have a constant effect on the variable under investigation.

The following time intervals were employed: Presentation rate—2·5 sec. (see above). Recall period—25 sec. Interval before next presentation—15 sec.

### *Subjects and Instructions*

Thirty-five subjects took part in the experiment (10 female and 25 male). They were all students drawn from several Departments of the University. No subject was aware of the purpose of the experiment or of the manipulation of the lists until this series of experiments had been completed.

Identical instructions were read to each subject. They were told that 12 two-digit numbers would be presented in succession and were asked to repeat these numbers and to learn them. The possibility of new items appearing was described in the following way: "After the first presentation and recall test, I will present a further series of 12 numbers. These may or may not be the same numbers." Subjects were informed that the experiment would comprise eight presentations and were encouraged to treat the eight presentations as one continuous task.

Before the experiment proper subjects were given one presentation and one recall test using 12 numbers which were not used later. These numbers were all double numbers or numbers including a zero. The purpose of this practice was to accustom them to the time intervals involved and to the verbal responses required of them.

### *Results and discussion*

The mean number of items correctly recalled per subject over the eight tests ("C-score") was 20·31 (S.D. = 5·67). C-score naturally does not include the ballast items. Mean C-score for each recall test is shown in Table I. It will be noted that C-score remains fairly constant over the eight recall tests. A "typical" learning curve is of course not to be expected in this experiment since the lists to be learned were changing continually.

As had been anticipated, subjects tended to recall immediately some or all of the ballast items in the order in which these had been presented. They then recalled isolated items from the remainder of the list. The consequences of this can be seen

TABLE I

MEAN C-SCORE FOR EACH RECALL TEST IN EXPERIMENT I

| *Recall tests* | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2·86 | 2·97 | 2·77 | 1·92 | 2·28 | 2·40 | 2·68 | 2·43 |

TABLE II

MEAN C-SCORE IN EXPERIMENT 1 FOR EACH SERIAL ORDER POSITION

| *Serial order position* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 2·51 | 2·74 | 2·46 | 2·49 | 2·11 | 2·63 | 2·57 | 2·80 | 4·09 | 4·46 | 5·89 | 7·60 |

Positions 9–12 are "ballast" items, explanation of which appears in the text.

in Table II, which shows the mean number of items correctly recalled in each serial order position. Although a serial order effect is very noticeable for the ballast items, variance due to serial order position of the first eight items is not so marked. To establish whether a bias in the proportion of items receiving each number of presentations which appeared in each serial order position could constitute an extraneous variable of the type discussed above, a series of $\chi^2$ tests was carried out. The expected frequency of items receiving each number of presentations in each serial order position was calculated from the total number of items receiving that number of presentations. For example, since 912 items received one presentation, 114 of these are expected to occur in each of the eight serial order positions if the distribution of items is unbiased. The $\chi^2$ values for each number of presentations are well below the significance level. It may therefore be concluded that the experimental control of this variable was satisfactory. It is also felt that the secondary control of employing ballast items was successful in increasing the presentation-recall interval (minimizing the influence of immediate memory), whilst preventing selective rehearsal in that interval.

The probability of an item being correctly recalled (the p(C) measure) is of the relative frequency type. For instance, 912 items were presented once, of which 295 were correctly recalled. The probability of correct recall after one presentation ("$p(C_1)$") is thus taken to be 295/912, which is 0·324. The relative figures are given in Table III.

It can be seen that p(C) does not increase with increasing number of presentations of an item. The rank correlation (*rho*) was calculated between $p(C_1)$, $p(C_2)$, etc., and the number of presentations for each subject. A positive correlation of this type indicates that the subject in question has derived benefit from repetition of items. The mean of the sample of correlation coefficients for the 35 subjects is not expected to differ significantly from zero if the group as a whole has derived no benefit from repeated presentation of items. This mean is, in fact, +0·03, which is not significantly above zero.

It may therefore be concluded that the group derived no benefit from repetition of items in this task. These results cannot be reconciled with an incremental account of learning, since this demands that p(C) increases with increasing numbers of presentations. The constancy of p(C) for the whole group despite increasing numbers of presentations of an item in this task suggests that the response-learning phase of paired associate learning may be explained by an all-or-none hypothesis.

TABLE III

THE RELATIONSHIP IN EXPERIMENT 1 BETWEEN C-SCORE AND THE NUMBER OF TIMES THAT AN ITEM HAS BEEN PRESENTED

| | *Successive presentations of a given item* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Items presented .. | 912 | 552 | 339 | 190 | 119 | 66 | 35 | 20 |
| Total C-score .. | 295 | 167 | 114 | 57 | 39 | 21 | 12 | 6 |
| p(C) .. .. .. | 0·324 | 0·303 | 0·336 | 0·300 | 0·328 | 0·318 | 0·343 | 0·300 |

p(C) is the probability of a response term being correctly emitted on a recall test.

It is, however, interesting to consider intragroup differences in the effect of repetition under these conditions. For 19 subjects a positive *rho* was obtained between $p(C_1)$, $p(C_2)$, etc. and number of presentations. The mean number of items correctly recalled over the eight trials for these 19 subjects was 22·53. On the other hand, the remaining 16 subjects (for whom a negative *rho* was obtained) correctly recalled a mean of only 17·69 items. This difference between the two groups of subjects is significant, ($U = 83$; $p < 0{\cdot}05$), and it is tempting to hypothesize that under these conditions faster learners benefit from repeated presentation of items, whereas slower learners do not. Further evidence for this hypothesis is clearly required before it can be accepted with any confidence.

## EXPERIMENT 2

This experiment was conducted to investigate the relationship between repeated presentation of items and the associative phase of paired associate learning. To measure the associative phase alone, it must be ensured that no response-learning is involved in the experimental task. That is to say that the response items must be the same throughout the experiment, and that the subject must be thoroughly familiar with these items.

### *Apparatus and Method*

A series of paired associate items was developed in which the stimulus terms were letters between D and Z inclusive or double letters between DD and ZZ inclusive. The letters I and II were omitted from this series. The response terms were the numbers 2 to 9, such that numbers 2, 4, 6, 8, were employed five times and the remaining numbers six times. The response terms were randomly assigned to the stimulus items with two constraints: (*a*) Double letters (e.g. DD) should not have the same response term as the corresponding single letter (e.g. D). (*b*) Double letters should not have as a response term a number twice that which is the response term for the corresponding single letter.

The stimulus and response terms were stencilled onto 3 in. by 5 in. plain cards. On the reverse face of each card the relevant stimulus item only was stencilled. This was placed to the left-hand side of the card as it would have been if the response item had been present.

Items were presented to the subject by placing the list in a rack such that all cards were standing up behind one another. The cards leant back slightly, and only the front one could be seen by the subject. This in turn was concealed by a plain white card until presentation of the list started. In this way, removal of one card exposed the next one to a subject. A list always comprised eight cards, and was arranged so that each of the numbers between 2 and 9 occurred once as a response term in every list. In this way, the response terms in each list were always known to the subject. The task did not involve learning or "integrating" responses, but rather it involved learning which response term was associated with which stimulus term. The task is therefore equivalent to the associative phase of paired associate learning. After each presentation of a list of eight cards there was a recall test during which the stimulus terms alone were presented. Order of presentation was varied throughout the experiment, which was terminated after eight presentations and eight recall tests.

The experimental procedure was similar to that used in Experiment 1. Each pair correctly associated on a recall test was replaced by another pair bearing the same response term. During presentations of a list the subject was required to repeat aloud both items of each pair as it was exposed, and during recall tests subjects had to repeat aloud the stimulus item and, if he could, the associated number. No knowledge of results was given. The proceedings were recorded on tape and later transcribed.

The time intervals employed were as follows:—Presentation of a pair—2 sec. Interval between pairs—nil. Before recall test—5 sec. Test presentation of stimulus item—5 sec. After recall test—10 sec. All times were controlled by a metronome.

### *Subjects and Instructions*

Thirty-four subjects took part in the experiment (four female and 30 male). All were students drawn from several Departments of the University. None was aware of the purpose of the experiment or of the manipulation of the cards until this series of experiments had been completed.

Identical instructions were read to each subject. These included details of the materials employed, the verbal responses required of him, the distinction between presentations and recall tests, and the length of the experiment. Subjects were asked to "learn the associations between the letters and the numbers," and it was emphasized that each of the numbers between 2 and 9 would occur in each series of eight cards, although not necessarily in that order. The possibility of new cards appearing was described in the same manner as in the instructions for Experiment 1.

Before the experiment commenced subjects were given one presentation and one recall test as practice. These employed those letters and double letters which were not used later, randomly assigned to the numbers 2 to 9.

### *Results and Discussion*

The mean number of items correctly associated by each subject ("C-score") was 21·03 (S.D. = 5·24). Mean C-score for each recall test is shown in Table IV. It will be seen from this Table that mean C-score on test 1 is markedly lower than on later tests, whilst C-score remains fairly constant between tests 2 and 8.

TABLE IV

MEAN C-SCORE FOR EACH RECALL TEST IN EXPERIMENT 2

| *Recall tests* | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1·94 | 2·71 | 2·59 | 2·62 | 2·68 | 3·00 | 2·85 | 2·65 |

It occasionally happened that a subject showed a marked predisposition to associate correctly a given response term. For instance, certain subjects would consistently associate correctly the response term "2" after very few presentations. The experimental design involved inserting a different stimulus term with "2" as its response every time this occurred. From time to time, therefore, no new "2's" were available for insertion into lists presented to these subjects. In this case a previously presented pair was inserted in order to hold constant the length of the list.

This item was omitted from the scoring. Since re-insertion of correctly associated items was only necessary on eleven occasions throughout the experiment, it is felt that this is a valid procedure. It may, however, be responsible for depressing very slightly the C-score on tests 7 and 8, since re-insertions occurred only on presentations 7 and 8.

The effect of repeated presentation of items on the probability of a response term being correctly associated is shown in Table V. The $p(C)$ measure is on the same basis as that used in Experiment 1. It can be seen that the probability of a correct association after one presentation ("$p(C_1)$") is lower than after greater numbers of presentations. $p(C)$ does not, however, increase regularly with increasing numbers of presentations, as would be predicted by an incremental hypothesis. The general trend of the $p(C)$ curve is to remain horizontal between $p(C_2)$ and $p(C_8)$.

TABLE V

THE RELATIONSHIP IN EXPERIMENT 2 BETWEEN C-SCORE AND THE NUMBER OF TIMES THAT AN ITEM HAS BEEN PRESENTED

| | *Successive presentations of a given item* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Items presented .. | 892 | 557 | 312 | 179 | 103 | 66 | 37 | 19 |
| Total C-score .. | 265 | 206 | 108 | 57 | 34 | 23 | 15 | 7 |
| p(C) .. .. .. | 0·297 | 0·370 | 0·346 | 0·318 | 0·330 | 0·349 | 0·405 | 0·368 |

$p(C)$ is the probability of a response term being correctly associated on a recall test.

It should be noted at this point that the number of items presented for the larger numbers of times (say, 7 and 8) is small. This is necessarily the case in an experiment of this kind. Therefore the values of $p(C_7)$ and $p(C_8)$ as shown in Table III cannot be taken to be reliable.

The dual-factor hypothesis requires that $p(C_1)$ is significantly lower than $p(C_{2+})$, but that there is no significant increase between $p(C_2)$ and $p(C_8)$. That the first of these requirements is met is evidenced by a $p$ of less than 0·001 for the difference between $p(C_1)$ and $p(C_{2+})$. (Wilcoxon matched-pairs signed-ranks test.) The trend between $p(C_2)$ and $p(C_8)$ was investigated by the same procedure as that employed in Experiment 1. The rank correlations between $p(C_2) - p(C_n)$ and the number of presentations for each subject were calculated and the significance of the difference between the mean of this sample and zero computed. In this case the mean is +0·08, which is not significantly above zero ($t = 1{\cdot}59$). These findings cannot be reconciled with either an all-or-none or an incremental account of subjects' performance on this learning task. It is felt that they suggest that the associative phase of paired associative learning may be accounted for by the dual factor hypothesis (Warr, 1963).

## CONCLUSIONS

It is concluded that the two hypotheses under test receive support from the results presented here. Insofar as it is possible to generalize from the present findings, the response-learning phase of paired-associate learning would appear to be an all-or-none occurrence and the associative phase would appear to be open to interpretation in terms of the dual factor hypothesis.

These results are also of interest in demonstrating that the effect of repetition of items is not the same for all types of learning. Further experiments are needed to investigate the importance of repetition in differing kinds of task. And, in the meantime, attempts to generalize from one set of results to other learning situations must be treated with considerable caution.

This research was supported by a grant from D.S.I.R. to the Department of Psychology, University of Sheffield for research into training procedures. The writer is grateful to Professor H. Kay and Mr. J. K. Clarkson for their constructive suggestions during the preparation of this paper.

## REFERENCES

CLARK, L. L., LANSFORD, T. G., and DALLENBACH, K. M. (1960). Repetition and associative learning. *Amer. J. Psychol.*, **73**, 22–40.

COOK, J. O., and KENDLER, T. S. (1956). A theoretical model to explain some paired associate learning data. In FINCH, G., and CAMERON, F. (Eds.), *Symposium on Air Force Human Engineering, Personnel and Training Research*, Publication 455. National Academy of Science—National Research Council, Washington D.C.

ESTES, W. K. (1960). Learning theory and the new "mental chemistry." *Psychol. Rev.*, **67**, 207–23.

FECHNER, G. T. (1860). *Elemente der Psychophysik*, **2**, 494.

HOROWITZ, L. M. (1961). Free recall and ordering of trigrams. *J. exp. Psychol.*, **62**, 51–7.

MCGUIRE, W. J. (1961). A multi-process model for paired associate learning. *J. exp. Psychol.*, **62**, 335–47.

MURDOCK, B. B. (1959). A reply to Battig, Gibson and Runquist. *Psychol. Rev.*, **66**, 345–6.

POSTMAN, L. (1961). Extra-experimental interference and the retention of words. *J. exp. Psychol.*, **61**, 97–110.

ROCK, I. (1957). The role of repetition in associative learning. *Amer. J. Psychol.*, **70**, 186–93.

ROCK, I., and HEIMER, W. (1959). Further evidence of one trial associative learning. *Amer. J. Psychol.*, **72**, 1–16.

TITCHENER, E. B. (1915). *A Beginner's Psychology*, 74.

UNDERWOOD, B. J., RUNQUIST, W. N., and SCHULZ, R. W. (1959). Response learning in paired associate lists as a function of intralist similarity. *J. exp. Psychol.*, **58**, 70–8.

UNDERWOOD, B. J., and SCHULZ, R. W. (1960). *Meaningfulness and Verbal Learning*. New York.

WARR, P. B. (1963). The effect of repeated presentation of items on paired associate learning. *Quart. J. exp. Psychol.*, **15**., 262–72.

*Manuscript received* 13*th November*, 1962.

# TEMPORAL STRUCTURE OF LETTER GROUPS AND SPAN OF PERCEPTION

BY

J. A. MICHON

*From the Institute for Perception RVO-TNO, Soesterberg, The Netherlands*

The span of perception for letter groups depends on number of letters presented, length of presentation and structure of the groups. The experiment reported varied the temporal structure of the groups, leaving the total number of letters constant. Groups of 12 letters were presented as a whole or in two or more successive "units." The total time of presentation was 1·5 sec. In the first experiment each unit was visible until the next appeared, in the second experiment units were visible only during ¼ sec., although intervals between successive units were kept constant.

The following conclusions emerged:

(*a*) The visual presence of units did not affect the reproduction for durations over 0·25 sec., except in the 12-letter presentations.

(*b*) 2 × 6 letters gave better results than either simultaneous presentation or other divisions; temporal separation was 0·75 sec.

(*c*) Higher order approximations to Dutch have more influence on 3 × 4, 2 × 6 and 1 × 12 letters than on 4 × 3 and 6 × 2 letters.

(*d*) A serial order effect exists: central units are reproduced less well than first and last units.

It is suggested that handling a fixed amount of information within a fixed period is limited on the one hand by confusion between simultaneous elements and on the other hand by the interaction between successive units presented too rapidly to allow for proper operation of immediate memory.

The difference between span of perception and span of memory is stressed.

## INTRODUCTION

The number of stimuli that can be perceived together in one single glance is closely related to the duration of the stimuli. If this time is below about 0·1 sec. the intensity of the stimulus plays an important role as well. Hunter and Sigler (1940) have demonstrated that in this case the law of Bunsen-Roscoe for visual impressions holds. For longer periods the so called "Span of Perception" for numerous classes of stimuli (e.g. dots, numbers, letters, nonsense syllables) is generally assumed to be rather near the "magical number 7" (Miller, 1956). Recently Fraisse and Battro (1960) have shown that the time of presentation is also important in longer intervals, up to 2·00 sec. Therefore Graham's (1951, p. 902) remark that ". . . for durations below about 3 sec. and with many classes of stimulus materials . . . the span of discrimination is little better for long durations than for short," seems to be unjustified, as an increase in reproduction of at least some 50 per cent. is reported for letters, between 0·25 sec. and 2·00 sec. by Fraisse and Battro.

Reproduction can be improved by extrinsic structure of the stimuli.* This has been found for numbers of dots and numerous other stimuli (cf. Woodworth and Schlosberg, 1954) and is probably valid for random letters as well, be the latter caused by gestalt principles, making some letters stand out more clearly, or by some *pons asinorum* imposed upon the groups by the subject, for instance by the formation of pseudo-words. It is clear anyway that "meaning" or even pronounceability increases the span enormously, because of the "gestalt structure" or "redundancy" of the new units: in this case the subject has no need to make up an "intrinsic" structure himself. Already Cattell and Wundt have mentioned the effect of the formation of larger units.

---

* Extrinsic structure (e.g. grouping) is opposed here to intrinsic structure (e.g. meaning).

Much of what is said thus far applies not only to simultaneous stimuli, but to successive stimuli as well. Flashes of light, auditory "pips" (provided the rate of presentation is too rapid for counting), as well as called out numbers, letters, etc., all give a span of some 6 or 7. Again the length of the interval is very important.

The relation between span and time of presentation, however, is not linear, but a monotonic increasing function, approaching an upper limit, both with simultaneous and with successive stimuli. This suggests that presentation of the same amount of information (e.g. a group of random letters) in a number of separate, successive presentations, may give a better result in reproduction than one simultaneous presentation in the same total time. If on the other hand the successive presentations follow too soon after each other, perception may be impaired, because units cannot be kept separated from one another any more by the organism. Thus some optimal structure is expected to emerge.

This will be tested in the experiment to be described. If the hypothesis is supported, some practical conclusions can be drawn. In industrial as well as in military tasks (e.g. controlling, coding, plotting) closely packed messages have to be handled by the operators. The density of the information determines what has been called the "perceptual load" of the task. If this information can be given an optimal temporal structure, either the perceptual load will be diminished somehow, or else the rate of presentation of information can be increased. This report describes two experiments.

## PROCEDURE

*Groups* of 12 letters were presented, either as a whole for 1½ sec., or in two successive *units* of 6 letters each for ¾ sec., or in 3 units of 4 letters (½ sec. each), 4 units of 3 letters (⅜ sec.) or 6 units of 2 letters (¼ sec. each). The complete sets of 12 letters will be called *groups* throughout the report, the subdivisions of 2, 3, 4 or 6 letters will be called *units*.

These objects were filmed in white letters on a black background on 8 mm. film. The film was projected on a white screen in a dimly lit room. One second before the first unit of a group of 12 letters appeared, a small white triangle was shown for ¾ sec. The film speed was 16 frames per sec. The angle under which the letters were presented was maximally 16° of visual angle in horizontal direction. This is not too far out of the field of clear perception (cf. Woodrow, 1938). The letter groups were generated in zero-, first- and second-order approximation of the Dutch language. This gives a total of 15 different groups of 12 letters. In order to eliminate as much as possible chance redundancy of letters, which invariably will occur when short sequences are generated, three parallel groups were constructed for each of the 15 conditions. Of these, the groups representing the median results were used in the final analysis. Thus each subject was presented with 45 groups of 12 letters. A sample of groups is given in Table I.

Subjects were instructed to reproduce as well as possible the 12 letters they were shown, paying attention to place and order, on a form on which the division of each group was indicated, like for instance the 3 × 4 letters: : . . . . : . . . . : . . . . : Prior to the actual groups 5 trials were given with more or less redundant groups, in order to make the subjects at ease in the, at first rather frustrating, situation. After each group presented, unlimited time was given to write down the answers. Each of two groups of 10 subjects was given a different series of 45 groups of 12 letters, in almost precisely the same way, to make cross-comparison of the results possible. The only difference was that in the second series the way of division of the group was announced by the experimenter. This was done after some comments of the subjects in the first series. No difference is to be expected from this procedure, because all subjects could see from their forms what kind of division was going to appear.

In a second experiment the same groups of letters were used as in the second series of the first experiment. This time the difference was the length of the actual presentation of

TABLE I

SOME EXAMPLES OF THE 12-LETTER GROUPS (2ND SERIES)

| *Approximation* | *Division* | |
|---|---|---|
| 0 | 1 × 12 | DLFJNGDGVYVQ |
| | 2 × 6 | JBLVAF GWUBZF |
| | 3 × 4 | LZDM JHAX PNXV |
| | 4 × 3 | WLJ JXH LVU WZB |
| | 6 × 2 | KZ GD AI UP ZL JI |
| 1 | 1 × 12 | GNEWZDEONEUX |
| | 2 × 6 | EBIENA THDOEI |
| | 3 × 4 | RDOE ILTE SNTI |
| | 4 × 3 | NTD EOB TLE WNE |
| | 6 × 2 | KA SG TL ED BE EZ |
| 2 | 1 × 12 | DOGEMODELMAK |
| | 2 × 6 | RUSTOR DAPPEC |
| | 3 × 4 | WOEN GAND ETRI |
| | 4 × 3 | PLA NSL IET ERD |
| | 6 × 2 | PR ID EN VR OD ST |

each unit. Instead of being visible until the appearance of the next unit (or until the end of 1½ sec.), every unit, whatever its number of letters, was presented for only ¼ sec. The remainder of the appropriate interval was left "black." With this procedure no visual examination of the letters was possible for more than ¼ sec. It will be noted that the condition of 6 units of 2 letters (6 × 2) is identical in both experiments, as each unit lasted exactly 1½/6 sec. = ¼ sec. in the first experiment.

The results were scored as follows:

> Each letter reproduced accurately and in the right place: 1 point.
>
> Each letter reproduced accurately and shifted one place: ½ point.
>
> Each group of n letters reproduced accurately and shifted one place: $n-1$ point.
>
> Maximal score: 12 points.

TABLE II

AN ANALYSIS OF VARIANCE ON THE DATA OF TWO INDEPENDENT REPLICATIONS OF EXPERIMENT I

| *Source of variance* | *d.f.* | *m.s.* | *F* | *p* |
|---|---|---|---|---|
| Replications .. .. .. .. | 1 | 0·30 | — | — |
| Residual between subjects .. .. | 18 | 23·27 | | |
| Approximations .. .. .. .. | 2 | 645·15 | 1040·56 | <0·001 |
| Ways of division .. .. .. .. | 4 | 206·90 | 333·71 | <0·001 |
| Approximations × Ways.. .. .. | 8 | 29·20 | 47·10 | <0·001 |
| Replications × Approximations .. | 2 | 0·75 | 1·20 | n.s. |
| Replications × Ways .. .. .. | 4 | 6·10 | 9·84 | <0·001 |
| Replications × Approximations × Ways | 8 | 6·71 | 10·82 | <0·001 |
| Residual within subjects .. .. .. | 252 | 0·62 | | |
| Total .. .. .. .. .. | 299 | | | |

In each of the two series of the first experiment and in the second experiment 10 subjects took part. All were male and had at least graduated from secondary school. About half the subjects were co-workers of the Institute for Perception, Soesterberg, The Netherlands, the others were ratings of the Royal Dutch Airforce.

## RESULTS

### *Experiment I*

First of all, a comparison was made between the two series of Experiment I, in order to see whether the results showed sufficient agreement to be combined and presented as they are in Figure 1. Therefore an analysis of variance was carried out on both sets of data. A summary of this analysis is given in Table II.

We see that by far the greatest part of the non-error variance is accounted for by the experimental variables (Approximations and Ways of division of the groups) and their interaction. On the whole *no* difference between the outcome of the two experiments is found, although some slight discrepancies between single groups exist. Together these account only for about 1 per cent. of the total variance, and they will therefore be neglected.

TABLE III

AVERAGE REPRODUCTION OUT OF 12 LETTERS FOR DIFFERENT APPROXIMATIONS

| | *Approximation* | | |
|---|---|---|---|
| | 0 | 1 | 2 |
| Average reproduced | 4·74 | 7·31 | 9·82 |

In Table III the average reproduction per group of 12 letters is given for the three approximations. The differences are significant at the 1 per cent. level. (The 1 per cent. critical value being equal to 0·98.) The influence of different ways of division on the reproduction is shown in Figure 1, for all three approximations separately.

From comparison it will be clear that the higher order approximations are more profitable for the longer units, 3 × 4, 2 × 6 and 1 × 12 letters being reproduced almost perfectly in the second order approximation. The effect is obviously caused by the higher intrinsic structure, which enables grouping to take place easier. The reason why the effect is much smaller in the 4 × 3 and 6 × 2 letter groups may be that the recording into new "chunks" (*sensu* Miller, 1956), which is made possible by the higher redundancy, is limited by the actual number of letters per unit, while on the other hand redundancy will be less recognizable in short units. It is significant in this respect that in no instance subjects have connected successive units and integrated them into larger "chunks," which invariably happens if successive strings of single letters are presented (as in spelling).

Secondly it was expected that in general some units would be perceived more clearly at the cost of others. To see whether this was actually the case, a further analysis was carried out on the results of the second series of 12-letter groups. The outcome is presented in Figure 2.

When more than two units are presented, the central units appear to be most affected, and the more a unit is surrounded by others, the more gravely it is affected. The effect seems to be equivalent to the serial order effect in experiments on the reproduction of simultaneously presented letter groups.

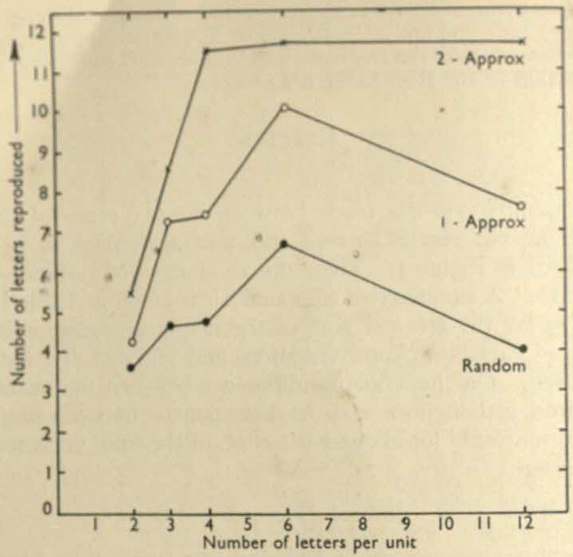


FIGURE 1

Number of letters out of 12 reproduced, as a function of approximation (intrinsic structure) and temporal (external) structure. Total time of presentation per group: 1½ sec. Time per unit = ⅛ sec. × number of letters per unit. 5 per cent. critical value of significant difference: $w_{0 \cdot 05} = 0 \cdot 85$ (20 subjects).

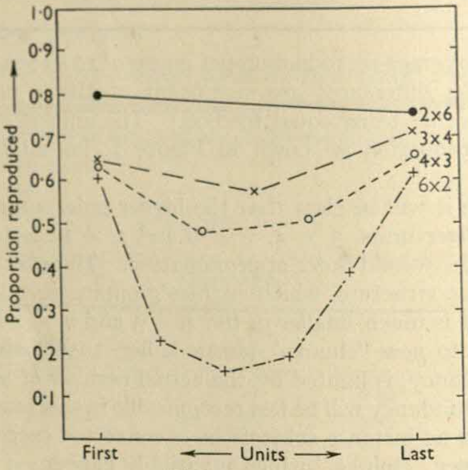


FIGURE 2

Proportion reproduced per unit as a function of number of units and of order of presentation. (Data of Experiment I, 2nd series; 10 subjects.)

A third conclusion emerging is that larger units give better reproduction, with a turning point at 2 × 6 letters. For this there are two possibilities. Either the time of presentation is too short for identification of sufficient letters in the case of 3 × 4, 4 × 3 and 6 × 2 letters but this appears to be unlikely under the conditions of this experiment, or the successive units interfere with each other because they are too close together in time. An incoming stimulus might be interfered with, when a second stimulus is presented within a period of 0·5 sec. or less, before it can be stored correctly in memory, or be handled in some other way (reaction). The reproduction of 2 × 6 is close to the maximal span, so that interference cannot have much influence beyond 0·5 to 0·7 sec. Thus the question arose whether reproduction will be affected when the longer units are present only for a short time while the interval between two successive units is not altered. Stated in other words: are the differences in reproduction as found in the present experiment determined by differences in visual duration? To answer this question Experiment II has been carried out.

*Experiment II*

This experiment differed in only one respect from the second series of the first experiment. All units were presented for ¼ sec. only, irrespective of their lengths. The rest of the appropriate intervals was left blank, so that the total time elapsing before reproduction was still 1½ sec. The results of Experiment II are presented in Figure 3. If we compare this diagram with Figure 1, it will be clear that they show close resemblance except for the reproduction of the 1 × 12 letter groups. Table IV demonstrates this point even more clearly. It summarizes an analysis or variance carried out on the results of the second series of Experiment I and of the present experiment, with and without taking the 1 × 12 letter groups into account.

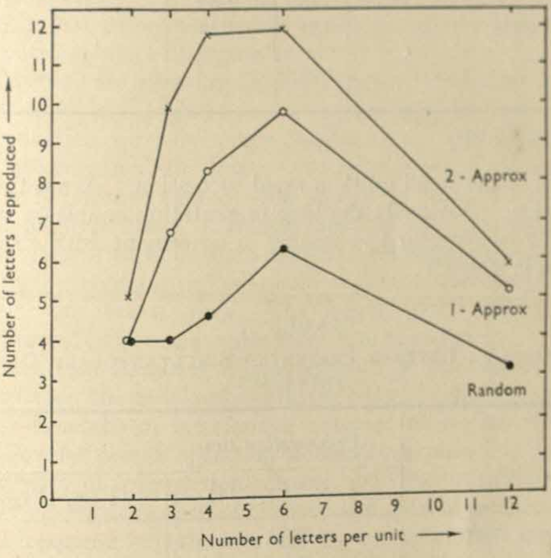


FIGURE 3

Average reproduction of letters per unit as a function of approximation and of number of letters per unit. Total time of presentation 1½ sec. Time per unit ¼ sec. irrespective of length. 5 per cent. critical value of significant differences: $w_{0 \cdot 05} = 1 \cdot 32$ (10 subjects).

The results of both experiments can be considered as identical, except for the 1 × 12 letter groups. This outcome permits the following conclusions: if we make the real presentation time shorter, then the amount reproduced is not affected in divisions from 2 × 6 to 6 × 2 letters. Obviously a presentation of only ¼ sec. per unit is sufficient in units up to 6 letters for the perception of all information which can be handled in the whole interval of 1½ sec. The rest of the interval is therefore only important for stabilization of the perceptual processes. This is the limiting factor in the reproduction of the small units in rapid succession. Again the 1 × 12 letters are the one exception. Obviously the long presentation enables a larger number of letters out of 12 to be perceived, a finding in agreement with Fraisse and Battro (1960), as is shown in Table V.

TABLE IV

ANALYSIS OF VARIANCE ON THE RESULTS OF EXPERIMENTS I (2ND SERIES) AND II. THE FIRST PART GIVES A SUMMARY OF THE COMPLETE ANALYSIS; IN THE SECOND HALF THE CONTRIBUTION OF 1 × 12 GROUPS IS TAKEN OUT

| | | Complete | | Partial | |
|---|---|---|---|---|---|
| *Source of variance* | *d.f.* | *m.s.* | *F* | *m.s.* | *F* |
| Duration units .. .. .. .. | 1 | 28·00 | 7·00** | 1·00 | — |
| Residual between subjects .. .. | 18 | 4·00 | | 3·94 | |
| Approximations.. .. .. .. | 2 | 561·50 | 398·23* | 445·00 | 386·96* |
| Ways of difference .. .. .. | 4 | 216·75 | 153·72* | 211·25 | 183·70* |
| Approximations × Ways .. .. | 8 | 27·50 | 19·50* | 26·38 | 22·94* |
| Duration × Approximations .. .. | 2 | 2·50 | 1·77 | 2·00 | 1·74 |
| Duration × Ways .. .. .. | 4 | 40·50 | 28·72* | 2·25 | 1·96 |
| Duration × Approximations × Ways | 8 | 7·87 | 5·85** | 1·63 | 1·41 |
| Residual within subjects .. .. | 252 | 1·41 | | 1·15 | |
| Total .. .. .. .. | 299 | | | | |

** $p < 0·01$ * $p < 0·001$.

TABLE V

REPRODUCTION OUT OF 12 LETTERS PRESENTED SIMULTANEOUSLY DURING DIFFERENT INTERVALS

| | Presentation time | |
|---|---|---|
| | *¼ sec. (Exp. II)* | *1½ sec. (Exp. I,2)* |
| Fraisse and Battro | 3·16 | 4·72 |
| Michon .. .. | 3·2 | 4·4 |

## DISCUSSION

The present experiment has answered to some extent the initial question. Under favourable circumstances the presentation in successive units of a quantity of information can improve the transmission, i.e. reproduction, in comparison to massed simultaneous presentation.

Gates (1916) has found that stimuli which do not exceed in number the span of perception, lead to the reproduction of more letters than larger numbers of stimuli. This fact has found confirmation by Fraisse and Battro (1960). These authors state that "les éléments se différencient moins quand ils sont trop nombreux pour être tous perçus" (p. 302). Crossman (1955) has introduced the concept of confusion between stimuli differing in some attributive dimension like length, brightness or colour (cf. Welford (1960) for a critical review). We can conceive of "confusion" as a measure of the interference between the perceptual processes, and therefore massed simultaneous presentation may be expressed in terms of Confusion Value, as was implicitly suggested by Fraisse and Battro (1960), because of interference between neighbouring letters (cf. also Woodrow, (1938)).

The results with respect to the 1 × 12 and 2 × 6 letter groups are in accordance with the expectations, 2 × 6 even being quite near the span generally reported as maximal (cf. Woodworth and Schlosberg, 1954). Although units of 4, 3 and 2 letters are easier to reproduce (a groups of 4 letters can be perceived without errors if the presentation time is 0·1 sec. or less), they lag behind the expected values in the present experiment. Data on interference of successive visual stimuli other than in simple reaction tasks are rare. Some observations of Sperling (1960) and Averbach and Sperling (1961) are relevant in this respect. They found that within a short period after visual presentation of letter groups subjects have more information available than they are able to reproduce—a datum which was reported introspectively by some of the subjects in the present experiment as well. Averbach and Sperling found also that successive stimuli can erase previous ones. This might offer a basis for our interference-effect with successive letter groups. One objection to this interpretation is that interference with earlier groups by later groups will not explain the serial position effect in Figure 2.

Some data in this field are given by Calabresi (1930). This author has found that two successive four letter groups have to be separated at least 0·75 sec. in order to give maximal reproduction, a result in agreement with our own findings. The closer successive stimuli are together, the heavier the interaction. Both the serial order effect found and the diminished reproduction at shorter unit-intervals are consistent with the one channel hypothesis of human information handling, and also with Broadbent's (1958) concept of immediate memory and rehearsal, as we can think of the smaller (fast) unit entering into the storage system before complete rehearsal of the previous unit(s) has taken place. The larger units will have had ample opportunity for rehearsal before new units enter into the system.

So in fact we find that handling a fixed amount of information within a fixed period is caught between the Scylla of confusion between simultaneously presented elements, and the Charybdis of interference between successive units, packed too closely in time to allow for proper operation of immediate memory. It is conceivable that the incompatibility of information content and discriminability (Crossman) is in fact the same as that between span of memory and span of perception (Fraisse and Florès (1956)). Subsequent experiments will be systematically controlled in this respect, by varying information content and discriminability independently.

Summarizing we reach the following conclusions:

(*a*) Temporal structuring will be advantageous when the complete message exceeds the amount of information which can be extracted from a simultaneous presentation (i.e. span of perception) without filling out the span of memory. When the latter has been filled, time should be allowed for further handling of its contents. Otherwise no further gain in reproduction can be expected.

(*b*) The advantages of temporal structuring tend to be lost if this leads to the parts succeeding on another by an interval of less than some 0·5 or 0·75 sec. At shorter intervals the proper functioning of short term memory is starting to be seriously affected.

## REFERENCES

AVERBACH, E., and SPERLING, G. (1961). Short term storage of information in vision. In CHERRY, C. (Ed.) *Information Theory*. London.

BROADBENT, D. E. (1958). *Perception and Communication*. London.

CALABRESI, R. (1930). *La Determinazione del Presente Psichico*. Florence. Cited by: FRAISSE P. (1957) *Psychologie du Temps*, Paris.

CROSSMAN, E. R. F. W. (1955). The measurement of discriminability. *Quart. J. exp. Psychol.*, **7**, 176–95.

FRAISSE, P., and BATTRO, A. M. (1960). L'évolution de la capacité d'appréhension en fonction du temps d'exposition. *Année Psychol.*, **60**, 295–307.

FRAISSE, P., and FLORÈS, C. (1956). Perception et fixation mnémonique. *Année Psychol.*, **56**, 1–11.

GATES, A. I. (1916). The mnemonic span for visual and auditory digits. *J. exp. Psychol.*, **1**, 393–403.

GRAHAM, C. H. (1951). Visual perception. In S. S. STEVENS (Ed.) *Handbook of Experimental Psychology*, New York.

HUNTER, W. S., and SIGLER, M. (1940). The span of visual discrimination as a function of time and intensity of stimulation. *J. exp. Psychol.*, **26**, 160–79.

MILLER, G. A. (1956). The magical number 7, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, **63**, 81–97.

SPERLING, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.*, **74**, No. 498.

WELFORD, A. T. (1960). The measurement of sensory-motor performance: survey and re-appraisal of twelve years' progress. *Ergonomics*, **3**, 189–230.

WOODROW, H. (1938). The effect of pattern upon simultaneous letter span. *Amer. J. Psychol.*, **51**, 83–96.

WOODWORTH, R. S., and SCHLOSBERG, H. (1954). *Experimental Psychology*, Ch. 4. London.

*Manuscript received* 17*th November*, 1962.

# EFFECTS OF VARIATION IN THE WARNING SIGNAL ON REACTION TIMES OF SEVERE SUBNORMALS

BY

BEATE HERMELIN

*From the Medical Research Council, Social Psychiatry Research Unit, Institute of Psychiatry, Maudsley Hospital, London, S.E.5*

Reaction times of normal and subnormal subjects were compared, when a warning stimulus varied in modality as well as in its temporal relation to the RT signal. When the warning stimulus was in a different modality from the RT signal, reaction times were shorter than when both were in the same modality. Mongols, in contrast to the other subjects, responded relatively faster to light than to sound.

The effect of varying the time interval between warning and signal was linearly related to response speed. The results are discussed in respect to the comparative alerting effects of temporal relations between stimuli and their different modalities.

## INTRODUCTION

Experiments on reaction time in which mentally deficient subjects are studied, consistently show a significant positive relationship between response speed and I.Q. Although there has been some analysis of this relationship (Annett, 1957; Berkson, 1960, I, II, III), the mechanisms involved are by no means clear.

In most reaction time procedures, a warning stimulus is presented to the subject prior to the onset of a signal to which he is asked to react. The warning seems to have the function of alerting the subject so that he is ready to respond to the signal when it appears.

There is some evidence, at any rate as far as physiological responses are concerned, that the mentally subnormal do not respond as much and for as long to short duration stimuli as normals. In three studies of the galvanic skin response the subnormal groups showed smaller amplitudes of response to brief visual and verbal stimuli, than did normal controls (Colman, 1931; 1959; Berkson, Hermelin and O'Connor, 1961). Alpha blocking after a light flash lasted for a shorter period in subnormals than in normals (Berkson, Hermelin and O'Connor, 1961). From these findings one may expect that a warning stimulus would alert subnormals less effectively and for a shorter period than normals, and that thereafter with increasing temporal separation between warning and RT signal, performance would deteriorate more rapidly in the subnormals.

Response speed in subnormals may also be affected by the respective modalities of warning stimulus and reaction time signals. These subjects tend to have impaired discrimination ability, and differences between stimuli usually need to be emphasized before they are perceived (Hermelin and O'Connor, 1960; O'Connor and Hermelin, 1961). One may therefore expect some degree of uncertainty and longer reaction times when warning and signal are in the same modality than when they differ.

## EXPERIMENT I

### *Subjects*

The subjects in the first experiment were 14 institutionalized severely retarded male patients and 14 members of the institution staff. The patients' I.Q.s ranged from 40–55 with a mean of 46. They were a clinically heterogeneous group. Chronological ages ranged from 18 to 30 years with no significant difference between the normals and the subnormals.

### *Apparatus*

The stimuli were mounted on a 36 in. by 24 in. vertical masonite screen, placed on a table in front of which the subject sat. A sound warning stimulus was a 400 c.p.s., 80 db., pure tone, emitted from a loudspeaker behind the screen. A warning light was seen through a circle 2 in. in diameter, placed 6 in. above the reaction time light signal, which in turn was seen through another 2-in. circle in the centre of the screen. Brightness of the warning light was matched in intensity with the warning sound according to the scale developed by Stevens and Galanter (1957). The brightness of the centre of the warning light was 1·2 log ft. lamberts and of the signal light 1·8 log ft. lamberts.

A second rectangle of masonite 12 in. by 15 in. lay in a frame on the table. A cadmium sulphide photosensitive cell (Mullard ORP 90) was mounted under a 0·2-in. hole in the centre of the masonite, and a light beam from a bulb mounted on top of the masonite shone through the hole on to the photocell. When the subject's finger was on the hole and this beam was interrupted, it was possible for the experimenter to begin a trial sequence. This sequence consisted of the warning stimulus presented for 0·25 sec., followed after a variable interval by the signal light. The signal light was turned off when the subject lifted his finger and the beam once again shone through the hole on to the photocell. The length of time the signal light was on was measured in 1/100 sec. units by a decade interval electronic clock and constituted the reaction time.

### *Procedure*

The subject was seated approximately 24 in. in front of the vertical screen in a darkened room. One experimenter remained in the room with the subject seated behind and to his left so that he could observe the subject without himself being seen. This first experimenter operated a switch which initiated trials and he also read the reaction times from the clock. The other experimenter, in an adjoining room, operated the dial which determined the warning-signal intervals. The following intervals were used, and are measured from the beginning of the warning stimulus: 0·3, 0·5, 1·0, 2·0, 4·0, 8·0 and 16·0 sec.

The reaction time response of lifting a finger was made to the white signal light in all cases. The warning was either the red light or the tone. Each type of warning was given on 70 successive reaction time trials so that there were 140 trials in all. Whether light or sound warning was presented in the first 70 trials or in the last was determined by a balanced design.

Each of the groups of 70 trials was divided into five blocks of 14 trials each. The different warning-signal intervals were presented in each block in a randomized order with the restriction that each interval was presented twice in each block. Thus within each block of 70 trials each interval was presented 10 times. Approximately 10 sec. separated trials. There were 30 sec. rest periods between the 14 trial blocks and a 3 min. rest between the first and second halves of the experiment.

A shilling was placed near the subject on the table at the beginning of the experiment. He was told that every time he responded fast enough, the experimenter would set aside a marble for him from a pile on the experimenter's desk, and that if he won "enough" marbles he could have the shilling. During the experiment, on any particular trial, the experimenter said "good" and set aside a marble if the subject responded faster than on the immediately previous trial. All the subjects were given the shilling after the experiment.

### *Results*

The reaction time data were analysed with regard to (*a*) modality of the warning stimulus; (*b*) order, i.e. the first vs. the second half of the experiment; (*c*) warning-signal interval, and (*d*) I.Q. The basic datum of the analysis was the median of the ten trials for each subject under each condition. A graph of the means of these medians is presented in Figure 1.

The reaction times are all rather long, even those of the normals. However, all subjects were untrained and were tested under conditions of long periods of uncertainty. Also, the stimulus light was of relatively low intensity. This may explain the unusually slow responses.

A separate analysis of modality and order effects was made for the normal and for the subnormal group. This showed, that with sound and light matched for intensity by Stevens and Galanter's scale (1957), for both groups the sound warning produced relatively faster reaction times than did the light warning (Normals: $F = 24{\cdot}70$, $p = 0{\cdot}001$, Subnormals: $F = 19{\cdot}59$, $p = 0{\cdot}001$). In addition, there was no significant order effect for the normals, but the subnormals responded more quickly in the first half of the experiment than in the second half ($F = 17{\cdot}12$, $p = 0{\cdot}005$).

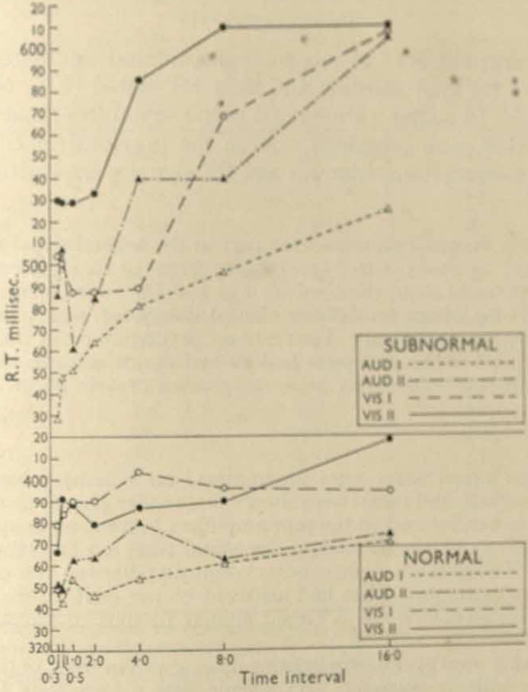


FIGURE 1
Median of trials

Analyses of the effects of warning-signal interval showed no significant interactions with Modality or Order. The data were therefore collapsed over Order and Modality and an analysis was performed to test the effects of Interval and of I.Q. For this analysis the medians were transformed to logarithms as the inclusion of both normal and subnormal raw data produced heterogeneous variances. The analysis demonstrated a significant interaction of Interval and I.Q. ($F = 5{\cdot}33$, $p < 0{\cdot}001$). Tests of simple effects showed that the normals were faster than the subnormals at all intervals, but this difference increased significantly with the progressive lengthening of the warning-signal interval.

These findings seemed to raise more questions than they answered. Firstly, the patients were an aetiologically and clinically heterogeneous group sharing intelligence levels between I.Q. 40 and I.Q. 55. Previous studies by Berkson (1960) and by Berkson, Hermelin and O'Connor (1961) had demonstrated differences between subgroups of subnormals in RT, as well as in extent and duration of physiological responses to short duration stimuli. The first question therefore was whether all or only some imbeciles had contributed to the effects of a longer forewarning period on RT.

Secondly the experiment raised the question whether the faster RTs in "sound-light" as compared with "light-light" conditions were due to some cross modality effects, such as had been found in previous studies (Hermelin and O'Connor, 1960; O'Connor and Hermelin, 1961). Alternatively they could be attributed to stable characteristics of the two modalities, and an auditory warning may have been more alerting than a visual. It was decided therefore to carry out another experiment.

## EXPERIMENT II

In this experiment the RTs of mongol, epileptic and undifferentiated imbeciles under conditions of (*a*) light warning and light RT signal (LL), (*b*) sound warning and light signal (SL), (*c*) sound warning and sound signal (SS), and (*d*) light warning and sound signal (LS) were compared. As in the previously mentioned study the time intervals between warning stimulus and RT signal were varied randomly.

### *Subjects*

Twelve severely subnormal patients took part in the second experiment. They were aged between 18 and 25 years with I.Q.s ranging from 35 to 48, mean I.Q. 43. Three subgroups of four patients each, matched for age and I.Q. were formed. The first contained four patients for whom no definite clinical diagnoses, apart from that of severe subnormality, had been established. Four mongol patients formed the second group and four epileptics the third. The epileptics had all had major seizures in the year prior to the experiment. All four were given twice daily doses of $\frac{1}{2}$ to $1\frac{1}{2}$ gr. of phenytoin and phenobarbitone.

### *Apparatus*

The warning and signal tones were programmed on separate channels on magnetic tape. The sound warning and signal were simply these tones played through loudspeakers. Sound operated relays, connected to the tape amplifiers, were used to operate warning and stimulus lights to the same programme. The sound relay on the stimulus channel also started the timer. This was a decade counter counting pulses at 100 per sec. To ensure that the warning and stimulus lights had achieved 80 per cent. of maximum brightness within 10 millisec. of switching on, a circuit similar to that of Carpenter and Leonard (1960) was used.

The visual warning was the uniform lighting up of a wall in front of the subject. White noise was used for auditory warning. RT signals were a pure tone and an illuminated circle 2 in. in diameter mounted on a vertical masonite screen. The intensity of sound and light was matched according to Stevens and Galanter's Scale (1957). Brightness of the warning light was 3·3 ft. lamberts and that of the signal light 63 ft. lamberts. The corresponding sound intensities were 72 and 80 db. respectively. Warning stimuli were presented for 0·25 sec., followed by a variable interval and the RT signal. Ten seconds after a response the next warning was given. The subject's response was the same as in Experiment I.

### *Procedure*

All subjects were tested under all conditions and a total of 200 RTs was recorded for each patient. The warning-RT intervals used were: 0·5, 2·0, 4·0, 8·0 and 16·0 sec. Each time interval occurred 10 times under each condition. The intervals were randomized over 50 trials with the restriction that each was presented twice in a block of 10 trials. Order of condition, i.e. light-light, sound-light, sound-sound and light-sound, was balanced between subjects according to a Latin square design. Trials were carried out in four sessions, one for each condition. At the beginning of each session the subject was told what to expect. Ten practice trials were then given. The same incentives as in the previous experiment were used.

### *Results*

RTs were on the whole shorter than those in the previous experiment. This difference was probably due to the higher intensities of the signals in the second study. The basis of analysis was the median of 10 trials for each subject under each condition and time interval. The means of these medians for groups, conditions and time intervals are given in Table I.

TABLE I

MEANS OF MEDIAN RT IN HUNDREDTHS OF A SECOND

| *Time intervals* | *Conditions* | *Epileptics* | *Imbeciles* | *Mongols* |
|---|---|---|---|---|
| 0·5 sec. | LL | 43 | 42 | 47 |
| | LS | 32 | 42 | 42 |
| | SS | 35 | 46 | 43 |
| | SL | 34 | 39 | 40 |
| 2 sec. | LL | 41 | 42 | 37 |
| | LS | 33 | 38 | 39 |
| | SS | 31 | 38 | 42 |
| | SL | 36 | 39 | 37 |
| 4 sec. | LL | 41 | 42 | 43 |
| | LS | 30 | 34 | 43 |
| | SS | 38 | 40 | 40 |
| | SL | 39 | 39 | 39 |
| 8 sec. | LL | 40 | 44 | 47 |
| | LS | 30 | 41 | 45 |
| | SS | 32 | 39 | 46 |
| | SL | 43 | 44 | 42 |
| 16 sec. | LL | 47 | 47 | 44 |
| | LS | 35 | 34 | 51 |
| | SS | 34 | 41 | 55 |
| | SL | 47 | 42 | 44 |

Separate analyses of variance were carried out for each group on the effect of time intervals and conditions. The results are shown in Table II.

TABLE II

RESULTS OF ANALYSES OF VARIANCE FOR TIME INTERVALS AND CONDITIONS

| | | *Epileptics* | | | *Other imbeciles* | | | *Mongols* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Source* | *d.f.* | *MS* | *VR* | *p* | *MS* | *VR* | *p* | *MS* | *VR* | *p* |
| Between intervals .. | 4 | 76·57 | | N.S. | 25·35 | | N.S. | 275·80 | 18·8 | 0·001 |
| Within squares: | | | | | | | | | | |
| Between subjects .. | 15 | 556·32 | | | 398·13 | | | 4302·51 | | |
| Between orders .. | 15 | 3105·82 | | | 5385·69 | | | 6586·26 | | |
| Between conditions .. | 3 | 457·58 | 7·72 | 0·001 | 240·52 | 10·05 | 0·001 | 127·55 | 8·7 | 0·001 |
| Intervals/conditions .. | 12 | 35·55 | | N.S. | 9·09 | | N.S. | 36·80 | 2·5 | 0·05 |
| Residual .. .. | 30 | 66·74 | | | 23·72 | | | 14·66 | | |
| Total .. .. .. | 79 | | | | | | | | | |

A highly significant difference between conditions was found in the epileptics. These subjects reacted relatively faster to sound than to light signals when these were matched for intensity, regardless of the modality of the warning stimulus. The RTs did not vary significantly for different warning-signal time intervals.

The undifferentiated imbeciles also reacted faster to sound than to light. In addition, RTs were significantly faster when a sound signal was preceded by a visual than by an auditory warning. The effect of modality of warning stimulus before a light signal approached but did not reach statistical significance. There were no significant effects of time intervals on RT.

The mongols, in contrast to the other two groups, reacted faster to the light than to the sound signal. The slowness in response to sound was particularly marked when both warning and RT signal were in the auditory modality. Visual warning before an auditory signal resulted in significantly faster RT. In visual RTs, faster responses were obtained when the light was preceded by a sound warning, than when light was used for the warning and for the signal. Thus mongols responded faster when warning stimulus and RT signal were in a different than when they were in the same modality.

The mongols also showed a significant effect of time intervals on RT. They reacted fastest when the warning preceded the signal by 2 or 4 sec. and slowest to a signal 8 or 16 sec. after the warning. With a short forewarning period of 0·5 sec. the RTs were significantly slower than after a 2 or 4 sec. and significantly faster than after an 8 or 16 sec. interval.

There was a significant interaction between conditions and time interval in the mongol group. This seems to be due to the fact that under conditions of visual warning and visual signal an interval of 0·5 sec. results in the longest RTs, while RTs after 16 sec. were considerably shorter than after an 8 sec. time interval.

The longest reaction times in these experiments had tended to be given under conditions of a warning-signal interval of 16 sec. Therefore RTs at 16 sec. interval formed the basis for a comparison between general response speed and decrease in RT with decreasing warning-signal intervals.

TABLE III

VALUES FOR REGRESSION LINES

x = RT 16 sec.
y = diff. RT 16 sec.–0·5 sec. =

$y' = 0{\cdot}65\,x - 19{\cdot}6$

$\sigma y{\cdot}x = 6{\cdot}5$

x = RT 16 sec.
y = diff. RT 16–2 sec. =

$y' = 0{\cdot}64\,x - 22{\cdot}3$

$\sigma y.x = 3{\cdot}9$

x = RT 16 sec.
y = diff. RT 16–4 sec.

$y' = 0{\cdot}63\,x - 22{\cdot}98$

$\sigma y.x = 4{\cdot}8$

x = RT 16 sec.
y = diff. RT 16 sec.–8 sec.

$y' = 0{\cdot}40\,x - 15$

$\sigma y.x = 0{\cdot}62$

Significant correlations between long RTs at 16 sec. time interval, and increasing speed with decreasing this interval were found for all subjects. Independent of diagnosis, varying the time intervals had about the same effect on all groups, provided they had similar reaction times at the longest warning-signal interval. Linear regression lines were calculated, relating RT at 16 sec. time intervals to RT 16–RT 0·5, 2, 4 and 8 sec. Values of their slopes and intercept values are given in Table III. The available normal data from the first experiment was included in this calculation.

As the regression lines relating RT 16 sec. to RT 16–0·5, 2 and 4 sec. did not differ significantly, the data for these time intervals was combined and a mean regression line calculated. This is illustrated in Figure 2.

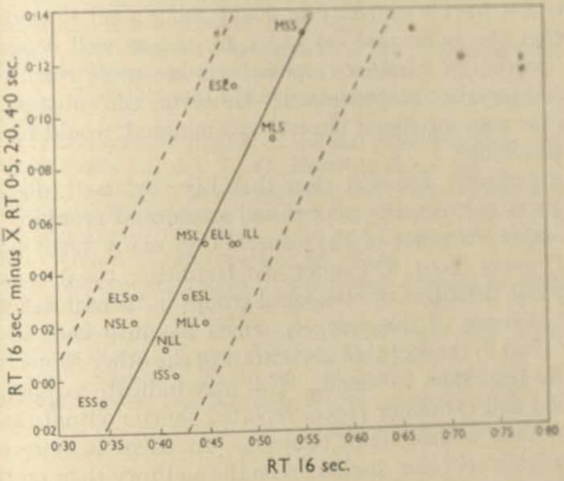


FIGURE 2

Regression line and standard error of y as estimated from x

As one would expect the effect of RT 16 sec. and the difference between RT 16–8 sec. was less marked though a significant relationship persisted.

## DISCUSSION

It might be argued that in order to draw conclusions about the relative alerting effects of different stimuli and warning-signal intervals, RTs under warning conditions should have been compared with those obtained when no warning was given. However, in order to equate procedures and preserve a similar degree of temporal uncertainty in warning and no warning conditions, such RT signals would have to be presented at irregular time intervals, covering a similar temporal interval range as those under warning conditions. Under such circumstances some signals would presumably be perceived before, and others after blocking of the alpha rhythm from the previous stimulus and response had ceased. In other words, some stimuli would have fallen within a period at which arousal from a previous signal still persisted, the latter thus acting as a warning. In other instances the arousal effect of the preceding stimulus and response would have subsided before the next one was presented. In those circumstances conditions of visual and auditory RTs without warning would have been similar to those obtained under light warning–light signal and sound warning–sound signal conditions. Therefore data of RTs taken under no warning conditions have been omitted.

In the first experiment, differences in RT of normals and subnormals increased when the time interval between warning stimulus and RT signal was lengthened.

The regression analysis of Experiment II suggests that this differential increment is not independent of the general level of a subject's RT. That is, the factors involved in the increasing separation between the groups may be similar to those differentiating the groups even on the shortest warning-signal intervals.

A previous experiment (Berkson, 1960) has suggested that normal and subnormal subjects differ, not so much in receptive or "judgmental" functions in a RT task, but in the speed of organizing and/or performing the response. The present experiments are consistent with the interpretation that relative to normals, mentally deficient patients also have a difficulty in maintaining a set to perform a particular response. Whether this is because responses are not as well organized in the subnormals at any interval or whether response sets are more readily interfered with over time (or both) remains undetermined. However, the substantial differences in group RT, when the warning-signal interval was minimal, would tend to support the first mentioned possibility.

The second experiment makes it clear that I.Q., by itself, does not adequately define the differences between the normal and subnormal groups. Mongols behave differently from other imbeciles of the same I.Q. in many tasks (see also Berkson, Hermelin and O'Connor, 1961; O'Connor and Hermelin, 1963). This points to the need for more precise definition of etiological groups in mental deficiency research.

The extreme slowness of the mongols, which has also been found by Berkson (1960), may be related to the fact that in contrast to the other subjects they responded relatively faster to light than to sound. This may indicate a generally low state of arousal. Venables and O'Connor (1959) reported shorter visual than auditory RTs in the slowest of their schizophrenic subjects. As there is more evidence for the effects of hypothalamic reticular discharges on the auditory than on the visual system (Bernhaut, Gellhorn and Rasmussen, 1952; Gellhorn, 1954; French, 1957; Lindsley, 1957) Venables has suggested that the relative performance of subjects in either auditory or visual modalities may be a measure of their level of total non-specific arousal.

The first experiment showed that visual RTs were shorter when the preceding warning stimulus was a sound, than when it was a light. From the second experiment it seems that when such effects occur they must be regarded as a function of having warning and reaction time stimuli in different modalities, rather than as a stable effect of the two modalities. If the latter were the case, the mongols, as they give shorter visual than auditory RTs, would be expected to react faster to a light when it was preceded by a visual than by an auditory warning. Actually the reverse is the case. Conversely the undifferentiated imbeciles should give the shortest RTs under sound-sound conditions. In fact responses are fastest when a light warning is given before a sound signal. However, the mechanisms which give rise to these cross modality effects are by no means clear, and further experiments are needed.

I am grateful to Dr. J. Gibson, Medical Superintendent, and Dr. B. W. Richards of St. Lawrence's Hospital, Caterham, for providing facilites for this experiment. The apparatus was designed and built by Mr. Eric Sayer, Senior Technician, Social Psychiatry Research Unit, Maudsley Hospital.

Dr. G. Berkson collaborated in the first study.

## REFERENCES

ANNETT, J. (1957). The information capacity of young mental defectives in an assembly task. *J. ment. Sci.*, **103**, 621–31.

BERKSON, G. (1960). An analysis of reaction time in normal and mentally deficient young men. I, II, III. *J. ment. defic. Res.*, **4**, 51–77.

BERKSON, G., HERMELIN, B. and O'CONNOR, N. (1961). Physiological responses of normals and institutionalized mental defectives to repeated stimuli. *J. ment. defic. Res.*, **5**, 30–9.

BERNHAUT, M., GELLHORN, E., and RASMUSSEN, A. T. (1952). Experimental contributions to problems of consciousness. *J. Neurophysiol.*, **16**, 21–35.

CARPENTER, A., and LEONARD, J. A. (1960). The use of small filament lamps as visual stimuli. *Quart. J. exp. Psychol.*, **12**, 249–51.

COLMAN, R. D. (1931). *The Psychogalvanic Reactions of Exceptional and Normal School Children.* Teachers College, Columbia University Centre Educ., 469.

COLMAN, R. D. (1959). The galvanic skin responses of mentally retarded and other children in England. *Amer. J. ment. Defic.*, **63**, 626–32.

FRENCH, J. B. (1957). Corticifugal connections with the reticular formation. In JASPER *et al.* (Eds.) *Reticular Formation of the Brain* (Chap. 24), London.

GELLHORN, E. (1954). Physiological processes related to consciousness and perception. *Brain*, **77**, 401–15.

HERMELIN, B., and O'CONNOR, N. (1960). Like and cross modality responses in normal and subnormal children. *Quart. J. exp. Psychol.*, **12**, 48–53.

LINDSLEY, D. B. (1957). The reticular system and perceptional discrimination. In JASPER *et al. Reticular Formation of the Brain.* (Chap. 25), London.

O'CONNOR, N., and HERMELIN, B. (1961). Like and cross modality recognition in subnormal children. *Quart. J. exp. Psychol.*, **13**, 48–52.

O'CONNOR, N., and HERMELIN, B. (1963). *Speech and Thought in Severe Subnormality.* Chapter VIII, London.

STEVENS, S. S., and GALANTER, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, **54**, 377–411.

VENABLES, P. H., and O'CONNOR, N. (1959). Reaction times to auditory and visual stimulation in schizophrenic and normal subjects. *Quart. J. exp. Psychol.*, **11**, 175–9.

*Manuscript received 21st December*, 1962.

# OVERTRAINING AND TRANSFER WITHIN AND BETWEEN DIMENSIONS IN THE RAT

BY

N. J. MACKINTOSH

*From the Institute of Experimental Psychology, University of Oxford*

Although reversal learning in the rat is normally assisted by overtraining, it seems that transfer to a new discrimination in a different stimulus dimension is not. This effect is confirmed in the present experiment, even though the stimuli of the second problem had been present (and necessarily attended to) in the first problem. This result is taken to show that the effects of overtraining cannot be accounted for by supposing that it strengthens an overt orienting response of the animal. A second group of rats was run to show that transfer to a second discrimination problem *within* the same dimension as the first is facilitated by overtraining. The important variable, therefore, that determines the direction of the effect of overtraining on transfer between problems, is the relationship between the stimulus dimensions concerned.

## INTRODUCTION

The learning of the reversal of a brightness discrimination by rats is facilitated if the animals have been overtrained on the original discrimination (Reid, 1953; Pubols, 1956). One way of deciding between various explanations of this finding is to see whether overtraining also facilitates transfer to an entirely different discrimination problem. Brookshire *et al.* (1961) found that overtraining in this situation made no difference to the speed of learning the second problem, while the present author (Mackintosh, 1962) found that it actually hindered such transfer learning. A possible explanation of this discrepancy will be discussed later; it is at any rate clear that overtraining does not assist transfer learning and this is incompatible with a number of theories of the overtraining-reversal effect (e.g. the application of Harlow's Error Factor theory (Harlow, 1959) and the suggestion that overtraining assists reversal because it reduces resistance to extinction (Capaldi and Stevenson, 1957)). An alternative explanation offered by both Reid and Pubols is that overtraining strengthens a "response of discriminating" which has the effect of isolating the relevant stimuli for the rat, thus ensuring that responses during reversal will still be controlled by these relevant stimuli. The predictions that follow from this account depend on further definition. Pubols seems to equate the response of discriminating with an overt orienting response: here the isolation of the relevant stimuli will mean that the rat is looking at the physical stimulus objects (and not at other parts of the apparatus); thus, provided the stimuli of the second discrimination problem are in the same place as the stimuli of the first, the strengthening of the orienting response with overtraining should presumably result in faster transfer learning. This prediction may not be absolutely rigorous, but becomes so if the two discrimination problems actually involve the same stimulus objects: this is achieved in the present experiment by training rats on an orientation discrimination between a square and a diamond: on half the trials both shapes are white, on the other half both are black. The test problem is a brightness discrimination between the black and the white square. Since the test stimuli were present during training, it clearly follows that the more firmly a response of orienting towards these stimuli has been established, the faster should the test discrimination be learnt.

However, it is possible that by a discriminating response is meant the isolation of a relevant stimulus *dimension*, and it is this interpretation that Brookshire *et al.* ascribe to Reid. Alternative formulations have been suggested by Goodwin and Lawrence (1955), and Sutherland (1959). If overtraining strengthens the isolation of a particular stimulus dimension, transfer to a new problem in a new dimension will be hindered by initial overtraining—regardless of the relationship between the physical stimulus objects in the two problems: thus even in the situation described above, the opposite prediction to that following from the orienting response theory will be made, and a test between the two theories becomes possible.

On the second theory, the feature of a reversal learning situation that accounts for the facilitating effect of overtraining, is that the initial problem and the reversal are two discriminations in the same stimulus dimension. There is in fact no evidence besides that coming from reversal learning to support the statement that transfer within a dimension is influenced by overtraining, and the second purpose of the present experiment is to investigate this point. Such an effect will not of course be found if the second problem is soluble simply on the basis of what has been learnt in the first: in other words the rats must begin the second problem without having preferences between the stimuli of this problem established by initial training. This can probably be achieved if the second problem is a simple brightness discrimination between a black and a white square, while the first problem is an absolute brightness discrimination between black *or* white versus grey: if both (or neither) black and white are rewarded, then the responses attached to the stimuli during initial training will have no relevance to the second problem, but isolation of the brightness dimension will be of relevance: the more this is learnt the faster the second discrimination will be solved.

The two groups so far described may also be compared with one another in their performance on the second problem, and this comparison provides another test between the two theories of the discriminating response. If it is stimulus dimensions that are isolated, then the group trained initially on the absolute brightness discrimination should (since for them both problems are soluble in terms of the same dimension) learn the second discrimination faster than the group trained initially on the squares and diamonds discrimination. The concept of the overt orienting response, however, provides no such grounds for differentiating the two groups, since they have both had the advantage of having been trained to look at the appropriate stimulus objects.

## METHOD

### *Experimental design*

There were two main experimental groups: Group I (16 subjects) was trained during Stage I to discriminate between a grey square and a black or white square; Group II (16 subjects) was trained during Stage I to discriminate between black or white squares and black or white diamonds (on any one trial both shapes were of the same brightness). Within each group half the subjects (subgroups I*a* and II*a*) were trained with the black and white squares positive, and the other half (subgroups I*b* and II*b*) with the black and white squares negative. All subjects were trained initially for 140 trials, and half received a further 140 overtraining trials. During Stage 2 all subjects learnt to discriminate between the black and the white squares, half being trained with black positive, half with white positive. Thus a 2 × 2 × 2 × 2 factorial design was used with two subjects per cell, and the factors were: dimension, positive stimulus and number of trials during Stage 1, and positive stimulus during Stage 2.

### *Subjects*

The subjects were 32 experimentally naïve male hooded rats from the colony maintained at the Institute of Experimental Psychology, Oxford. They were 4–6 months old at the beginning of the experiment.

### *Apparatus and pretraining*

A modified Lashley jumping stand was used. The jumping platform was Y-shaped and each arm faced one of the two stimulus windows; these were set at an angle of 130 degrees and separated by a partition. Landing platforms were fixed below each window onto which the subject jumped when making a response. The stimulus doors (15 cm.$^2$) were made of brown hardboard, and the stimuli of opaque black, white, or grey perspex, ¼ in. thick. All stimuli were squares of side 6·5 cm. and were glued centrally to the doors.

The subjects were put on a 1½ hr. feeding schedule, and pretrained on the apparatus for 7 days, for 2 days being given their entire daily food ration on the feeding platform behind the windows, and allowed to explore; for the remaining 5 days being trained to jump gradually increasing distances, at first to open windows, and finally to blank hardboard doors. For further details of the apparatus and pretraining procedures, *v.* Mackintosh (1963*a*).

### *Training*

*Stage* 1. Subjects were run 10 trials a day for 14 days by a modified correction procedure. If they jumped to the correct door, they could push it open and run through to the food bowl behind, where they were allowed to eat for 20 sec. After an incorrect response, they found the door locked and were detained on the landing platform for 10 sec. They were then replaced on the jumping platform and allowed to jump again. A trial continued until a correct response was made. No punishment was given for incorrect responses. The inter-trial interval, during which they were kept in individual detention cages, was 6–8 min. They were fed for 1½ hr. in their home cages after their daily trials. The positive and negative stimuli were changed from side to side according to selected random orders. A second random order determined whether black or white stimuli should be used on any one trial (for Group I, whether the grey square should be paired with the black or the white square, and for Group II whether both square and diamond should be black or white). Within each day's trials the positive and negative appeared equally often on either side, and black and white stimuli were used five times each.

*Overtraining*. Eight subjects in each group were given 140 overtraining trials, the procedure being exactly the same as that described above.

*Stage* 2. The procedure was the same as for Stage 1, except that a non-correction method of training was employed. After an incorrect response, subjects were left on the landing platform for 10 sec. and returned to their detention cage. They were trained to a criterion of 18 correct responses over any 20 trials, with the last 10 all correct.

## RESULTS

The results are shown in Table I. For simplicity, the results for subjects trained in opposite directions (during both Stage 1 and 2) are pooled together, although there were certain differences between various subgroups. These are discussed below.

TABLE I

MEAN SCORES DURING STAGE 1 AND STAGE 2

| *Groups* | *No. of trials during Stage* 1 | *Stage* 1 | | *Stage* 2 |
|---|---|---|---|---|
| | | *Trials to Criterion* | *Total errors during trials* 1–140 | *Trials to Criterion* |
| I | 140 | 118·7 | 52·7 | 45·6 |
| | 280 | 118·7 | 58·5 | 39·7 |
| II | 140 | 128·7 | 63·7 | 61·2 |
| | 280 | 128·7 | 67·1 | 80·7 |

### *Stage 1*

Two measures of performance during Stage 1 were computed: the number of trials to reach a criterion of 80 per cent. correct responses over 2 days' trials; and the total number of errors (both initial and repetitive) made over the 140 trials. The 80 per cent. criterion was reached by all subjects in both groups, and on this measure over-trained and non-overtrained subgroups in each main group were matched exactly; on the errors measure they were matched very closely. Breakdown of the scores of subgroups trained in opposite directions showed that in Group I, subjects trained with black/white positive (Group I*a*) made fewer errors than subjects trained with grey positive (Group I*b*): the mean error scores were 56·1 and 61·1 respectively: $(t = 2{\cdot}8,\ d.f. = 14,\ p < 0{\cdot}01)$. In Group II subjects trained with squares positive (Group II*a*) made more errors than those trained with diamonds positive (Group II*b*) —mean scores were 67·5 and 62·1 respectively, but this difference was not significant: $(t = 0{\cdot}9,\ d.f. = 14,\ p > 0{\cdot}10)$.

The use of a correction training procedure automatically equalises the number of rewarded responses to the test stimuli for subgroups I*a* and II*a* and the number of rewarded responses away from the test stimuli for subgroups I*b* and II*b*; and since the error scores of these latter subgroups were closely matched $(t = 0{\cdot}2,\ d.f. = 14,\ p = > 0{\cdot}10)$, the number of non-rewarded responses to the test stimuli was nearly the same for both. Thus both Group I and II had equivalent experience with the test stimuli during Stage I.

The overtrained subjects were performing only slightly more accurately on the last two days of overtraining (86·2 per cent. correct) than on the first 2 days (84·4 per cent. correct); a paired replicate *t*-test of the difference in error scores was not significant $(t = 1{\cdot}4,\ d.f. = 15,\ p > 0{\cdot}10)$.

### *Stage 2*

The measure of learning given in Table I is trials to criterion. An analysis of variance on all subjects combined showed a highly significant effect of the dimension of training during Stage 1 $(F = 58{\cdot}9,\ d.f. = 1{,}20,\ p < 0{\cdot}001)$, and a significant interaction between training dimensions and overtraining $(F = 11{\cdot}8,\ d.f. = 1{,}20,\ p < 0{\cdot}01)$. Analyses on pairs of subgroups gave the following results: the difference between Group I and Group II was significant even when only non-overtrained subjects were considered $(F = 8{\cdot}8,\ d.f. = 1{,}12,\ p < 0{\cdot}05)$; overtraining had a significant effect on Group II $(F = 8{\cdot}0,\ d.f. = 1{,}10,\ p < 0{\cdot}05)$, but did not have a significant effect on Group I $(F = 2{\cdot}4,\ d.f. = 1{,}11,\ p > 0{\cdot}10)$. Thus subjects trained on the absolute brightness discrimination in Stage 1 learnt the test brightness discrimination faster than subjects trained on the shape discrimination: overtraining slightly facilitated the former, but significantly hindered the latter.

*Direct transfer*. It is important for the conclusions to be drawn from this experiment that no subject should begin the Stage 2 discrimination with preferences between the black and white squares already established by initial training. By training Group I on an absolute brightness discrimination, it was hoped that no such preference would be established. Similarly it was hoped that the square-diamond discrimination for Group II would involve no unsuspected brightness differences. One cannot say a priori whether this will be the case: hence the importance of using a factorial design, since this permits testing for the presence of this sort of direct transfer by the interaction terms in the analyses of variance. If direct transfer occurred then two interactions might be expected to be significant: the first-order interaction between the positive stimuli of Stage 1 and Stage 2; and (since if direct transfer occurs it should be magnified by overtraining) the second-order interaction with overtraining. In fact in the two analyses carried out on the Stage 2 scores of Group I and Group II separately, neither of these interaction terms were significant. For a demonstration, however, that such unexpected direct transfer can occur in this type of experiment, *v.* Mackintosh (1963*b*).

## Discussion

### *Stage* 1

An incidental finding of some interest concerns the development of position habits during the initial stages of training. On days 1 and 2, no subject showed a marked preference for either position, and the average number of responses to the preferred position was 12·7 out of 20. On days 3 and 4, the average was 15·4. This increase in the strength of position habits was significant beyond the 0·1 per cent. level of confidence, as measured by a *t*-test for paired replicates ($t = 5{\cdot}5$, $d.f. = 31$). The fact that rats with initially rather weak position habits develop stronger position habits during the course of the next few days' training, supports the idea that position habits are systematic attempts at solution of the discrimination problem, or "hypotheses," as suggested by Krechevsky (1932). Spence (1936) ingeniously showed that it was possible to account for this development of position habits within the framework of his continuity theory; but this account only holds for a non-correction method of training. As Spence (1940) has pointed out, on his theory a correction method of training should lead to a rapid reduction in position habits. The present results, of course, were obtained using a correction procedure, and are therefore inexplicable in terms of Spence's theory.

### *Stage* 2

The results for Stage 2 tell strongly against the interpretation of the response of discriminating as an overt orientation of the animal. Despite the fact that Group II must have learnt during Stage 1 to orientate towards the stimuli later to be used in Stage 2, overtraining on Stage 1 significantly hindered Stage 2 learning. This result agrees with an earlier finding (Mackintosh, 1962). In the case of Group I, the effect of overtraining was slightly to assist Stage 2 learning. It is clear, therefore, that the effects of overtraining depend upon the relationship between the dimensions of training and test problems, and in order to account for the results, the response of discriminating must be interpreted as the act of isolating the relevant stimulus dimension.

Although Group I was benefited by overtraining, the effect was small and not significant. Within the terms of, say, Sutherland's theory, it is not hard to see why this is so. The explanation for the superiority of the overtrained group is that they have learnt to switch in the relevant analyser (i.e. have learnt to isolate the relevant dimension) more firmly. Unless, however, the non-overtrained group switch out the relevant analyser upon transfer to Stage 2, they will still learn as fast as the overtrained group. Upon transfer, the percentage of trials on which reward is received will drop from about 100 to 50 per cent., and this may only cause non-overtrained animals to switch out the analyser incompletely and temporarily. The greater effect of overtraining found in reversal experiments will be because upon reversal there is a 100 per cent. drop in reward, and this will cause the non-overtrained animals to switch out the analyser completely.

If overtraining causes an animal to switch in the relevant analyser more firmly, one might expect that in Stage 1 subjects should perform increasingly accurately during the course of overtraining. There is in fact a slight improvement, but it is not significant. The asymptote of learning had in effect been reached before overtraining began. This can only mean that subjects continued to make errors not because they were failing to switch in the relevant analyser, but because they were failing to detect differences between positive and negative even with the analyser switched in. Both Stage 1 discriminations were difficult; it must be assumed that the differences between the stimuli were small enough sometimes to be obscured by noise. Both at the beginning and at the end of overtraining, subjects were switching in the relevant analyser on all trials; at neither time would this lead to perfect performance.

The final comparison of interest lies in the Stage 2 performance of non-overtrained subjects in Group I and Group II. Group I was significantly superior; this would be expected on Sutherland's theory, because for Group I both problems involve the same stimulus dimension, while for Group II the transfer from Stage 1 to 2 involves a change of dimension. Since for neither group does the transfer involve a change of stimulus objects, the difference cannot be explained in terms of the orienting response.

There remains the factual discrepancy between the author's results and those reported by Brookshire *et al* (1961). The latter failed to find any effect of overtraining on transfer to a new problem, while both here and previously the author has found overtraining to hinder transfer learning. Brookshire *et al.* trained their rats in a cross-maze; they were run from either of two start-arms, and learnt one of two problems: "place learning" involved always selecting a particular goal-arm, regardless of whether this meant turning right or left at the choice-point; "response learning" involved always turning, say, right at the choice-point regardless of which goal-arm such a right turn led to. They were then transferred to the other problem with or without overtraining. In order that this experiment should be a test of the point at issue, it is important that neither problem should be soluble in terms of the other; but it is by no means clear that this is the case. It is quite possible that the response learning problem should have been solved as a conditional place discrimination: the subjects might have learned to select one goal-arm when run from one start-arm, and the other when run from the other. Indeed there is some evidence that this is the normal basis for solution of a response learning problem (Ritchie *et al.*, 1951). In this case the two problems would not have been independent, and the experiment not a test of transfer between dimensions.

This work forms part of a project on "Stimulus Analysing Mechanisms": the project is financed jointly by the American Office of Naval Research (Contract N62558–2453) and the Nuffield Foundation.

## REFERENCES

BROOKSHIRE, K. H., WARREN, J. M., and BALL, C. G. (1961). Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Physiol.*, **54**, 98–102.

CAPALDI, E. J., and STEVENSON, H. W. (1957). Response reversal following different amounts of overtraining. *J. comp. physiol. Psychol.*, **50**, 195–8.

GOODWIN, W. R., and LAWRENCE, D. H. (1955). The functional independence of two discrimination habits associated with a constant stimulus situation. *J. comp. physiol. Psychol.*, **48**, 437–43.

HARLOW, H. F. (1959). Learning set and error factor theory. In S. KOCH (Ed.). *Psychology: a study of a science.* Study I, Vol. 2, New York, p. 492–537.

MACKINTOSH, N. J. (1962). The effects of overtraining on a reversal and a non-reversal shift. *J. comp. physiol. Psychol.*, **55**, 555–59.

MACKINTOSH, N. J. (1963*a*). The effect of irrelevant cues on reversal learning in the rat. *Brit. J. Psychol*, **54**, 127–34.

MACKINTOSH, N. J. (1963*b*). Direct transfer from a horizontal-vertical discrimination to a brightness discrimination in the rat. *Quart. J. exp. Psychol.*, **15**, 212–3.

PUBOLS, B. H., Jr. (1956). The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, **48**, 243–8.

REID, L. S. (1953). The development of noncontinuity behaviour through continuity learning. *J. exp. Psychol.*, **46**, 107–12.

RITCHIE, B. F., HAY, A., and HARE, R. (1951). Studies in spatial learning: IX. A dispositional analysis of response-performance. *J. comp. physiol. Psychol.*, **44**, 442–9.

SPENCE, K. W. (1936). The nature of discrimination learning in animals. *Psychol. Rev.*, **43**, 427–49.

SPENCE, K. W. (1940). Continuous versus non-continuous interpretations of discrimination learning. *Psychol. Rev.*, **47**, 271–88.

SUTHERLAND, N. S. (1959). Stimulus analysing mechanisms. In *Proceedings of a Symposium on the Mechanisation of Thought Processes*. Vol. II, London, p. 575–609.

*Manuscript received 3rd December, 1962*

# PERIPHERAL VIEWING AND SIMULTANEOUS CONTRAST

BY

PETER BURGH

*From the Psychological Laboratory, University of Cambridge*

An experiment was done to find the effect on simultaneous brightness contrast of viewing the test patches foveally or peripherally, at a distance of 3° or 6° from the fovea. It was found that contrast was greater in the periphery. It was also found that contrast increased with prolonged viewing of the display. A further experiment showed that blurring the test patches produced an increase in contrast.

## INTRODUCTION

There is a considerable body of evidence that the receptive fields of the foveal region differ both in size and in degree of inhibitory action from those of the peripheral retina (Wiesel, 1960; Barlow, FitzHugh and Kuffler, 1957).

If it is accepted that there is at least a strong retinal component in simultaneous contrast (Asher, 1950), then it might be expected that lateral inhibition played some part in simultaneous contrast phenomena, and that these phenomena also would differ in foveal and peripheral parts of the retina. However it is not clear what should be expected in the field of human perception, on the basis of the electrophysiological findings. It may be that this is a field in which it could be useful to establish psychological or psychophysical facts and from these to make physiologically testable predictions.

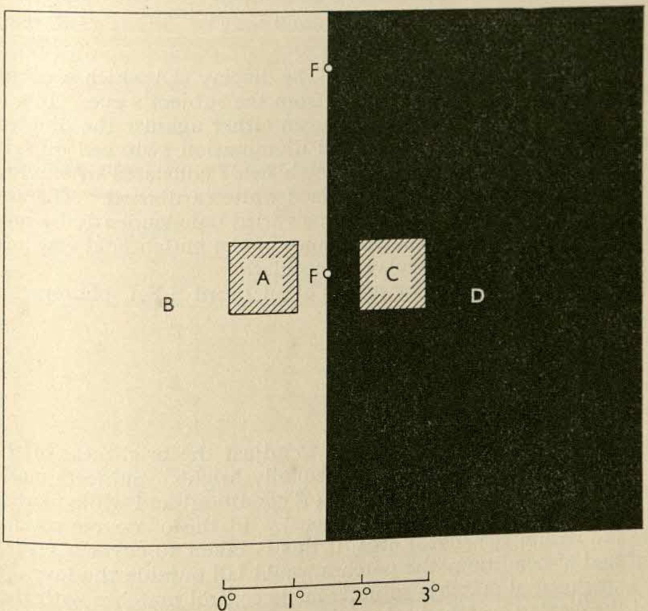


FIGURE 1

The 6° fixation point is not shown.

Preliminary investigations suggested that there was considerable variability in the judgements of any observer upon different occasions, and a further aim of this experiment was to compare interpersonal and intrapersonal variation in order to discover what in fact was the degree of correlation between the judgements of one subject in different experimental sessions.

## EXPERIMENT I

### *Display and Apparatus*

The display to the subject was of the conventional type as illustrated in Figure 1. Area (A), the test field, was of lower luminance than its "white" surround (B), while the Match field (C) was of higher luminance than its "black" surround (D).

Areas (A) and (C) had sides which subtended 1° of visual angle at the subject's eye, and were also separated by 1°. The whole display measured approximately 25° of visual angle in width, and 20° in height. Three fixation points (F) were provided: one was directly between the patches, and the others were placed towards the top edge of the display at distances subtending 3° and 6° respectively from the central point.

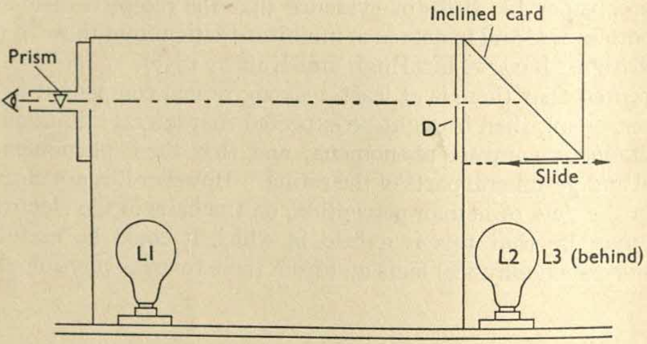


FIGURE 2

The apparatus is illustrated in Figure 2. The display (D) which was cut out of cardboard, was placed vertically at about 60 cm. from the subject's eye. It was illuminated by lamp L1, so that the patches could be shown either against the divided black-white surround as illustrated, or with the surround illumination switched off, giving in effect a uniform black surround. The test and match fields consisted of windows cut in the display behind which were two pieces of inclined white cardboard. The amount of light falling on these from lamps L2 and L3 could be varied independently by means of sliding shutters. The shutter controlling the luminance of the match field was adjusted by the subject. The subject viewed the display monocularly.

The luminances were measured by means of an Ilford S.E.I. photometer and were as follows:

White surround (B) 26·9 ft. L.
Black surround (D) 1·0 ft. L.
Test field (A) 18·2 ft. L.
Match field (C) from 0·8–33·9 ft. L.

In preliminary trials subjects were asked to adjust the brightness of the match field until the two patches (C) and (A) appeared equally bright. Subjects made judgements both while fixating the central fixation point (0° condition) and while fixating each of the peripheral fixation points (3° and 6° conditions). In the 0° degree condition both the patches would fall within the foveal area, if this is taken to cover a visual angle of 3°; while in the 3° and 6° conditions the patches would fall outside the foveal area. Before and after the experimental matches subjects made control matches with the illumination to the front of the display switched off, so that both patches were seen against a uniform dark surround. This was done in order to compare the scatter of readings taken with and without the effect of contrast.

The results of these preliminary trials suggested that the effect of contrast was greater under the 3° and 6° conditions than at 0°. However this result was unsatisfactory for the following reasons:

(1) It was found that the control matches with uniform surround gave different means before and after the experimental matches. After the experimental matches the match field was set to a lower luminance than the test field, as though the surround were still illuminated. This suggested a cumulative adaptation effect lasting throughout the session. In order to overcome this a prism was added to the eyepiece so that the field could be reversed from left to right between each trial. This it was thought would prevent the selective adaptation of one part of the retina over a relatively long period.

(2) Several subjects reported that the effect of contrast appeared to increase during each trial. That is to say, that with continued fixation of the display the test field seemed to become darker and the match field lighter than at the beginning of the trial. In order to discover whether this was really the case the instructions were altered. Subjects were asked to make a match quickly, in 5 sec., but after this to continue to fixate the display for 30 sec., adjusting the brightness if necessary to maintain the match.

(3) In order to avoid kinaesthetic or tactile cues from the position of the slide, the subject was asked to move the slide by means of a knob which, before each match, was moved in haphazard order to one of three different positions on the slide.

Ten subjects, students or research workers in the Psychological Laboratory, Cambridge, were tested with these changes in apparatus and instructions. As before, each session began with four control matches with the uniform dark surround, followed by eight experimental matches on each fixation point in the contrast condition, and finishing with four more control matches. The fixation points were taken in different orders by different subjects; each subject served for two sessions, on different days, and in each session a subject was presented with the three fixation points in a different order. The sessions lasted approximately ½ hr.

### *Results*

The results are shown in Tables I, II and III. They are given in terms of Degree of Contrast: that is to say, the ratio of the luminances of the match field and the test field at the matching point, so that for example 3·1 means: "The brightness of the patches was judged equal when the test field (on the white surround) was 3·1 times brighter than the match field."

TABLE I

MEDIAN MATCHING POINT, DEGREE OF CONTRAST (MATCH/TEST FIELDS)

| *Subject* | *5 sec.* | | | | *30 sec.* | | | | *Total average* |
|---|---|---|---|---|---|---|---|---|---|
| | 0° | 3° | 6° | *Average* | 0° | 3° | 6° | *Average* | |
| 1 | 2·2 | 3·4 | 3·7 | 3·1 | 2·5 | 5·1 | 4·2 | 3·8 | 3·5 |
| 2 | 1·95 | 2·3 | 2·35 | 2·2 | 1·85 | 3·1 | 3·3 | 2·7 | 2·5 |
| 3 | 2·85 | 4·2 | 4·4 | 3·8 | 3·3 | 5·25 | 5·6 | 4·7 | 4·3 |
| 4 | 3·1 | 3·4 | 4·5 | 3·7 | 3·5 | 4·8 | 6·8 | 5·0 | 4·3 |
| 5 | 3·4 | 3·55 | 5·4 | 4·1 | 4·35 | 5·9 | 6·6 | 5·6 | 4·9 |
| 6 | 2·25 | 3·85 | 5·0 | 3·7 | 2·75 | 5·25 | 5·9 | 4·6 | 4·2 |
| 7 | 3·8 | 4·0 | 5·6 | 4·5 | 4·0 | 5·5 | 7·0 | 5·5 | 5·0 |
| 8 | 2·8 | 3·1 | 3·25 | 3·05 | 3·1 | 4·5 | 4·0 | 3·9 | 3·5 |
| 9 | 2·85 | 3·5 | 4·8 | 3·7 | 2·85 | 3·4 | 4·7 | 3·65 | 3·7 |
| 10 | 3·0 | 3·7 | 4·0 | 3·6 | 3·3 | 4·8 | 5·0 | 4·4 | 4·0 |
| Average | 2·8 | 3·5 | 4·3 | 3·5 | 3·1 | 4·8 | 5·3 | 4·4 | 4·0 |

(1) Clearly peripheral fixation gives more contrast than central, and the 6° condition more than 3°. Also matches made at 30 sec. give more contrast than those

made at 5 sec. The fixation differences are significant for all subjects and the temporal differences for eight out of the 10 ($p < 0\cdot01$), when tested by a distribution free analysis of variance (Wilson, 1956). The use of non-parametric statistics seemed advisable in view of the lack of grounds for assuming normality of the distributions.

(2) The scatter (expressed by the Interquartile Range as a percentage of the Median, $\frac{\text{IQR} \times 100}{\text{Md}}$) was compared for the various conditions.

There were no significant differences in scatter between the experimental conditions. But the scatter for each experimental condition was significantly greater than that for the control condition, ($p < 0\cdot01$), by the Wilcoxon Test (Siegel, 1956).

TABLE II

SCATTER. INTERQUARTILE RANGE AS PERCENTAGE OF MEDIAN

| *Subject* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | *Average* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Experimental .. .. .. | 37 | 32 | 27 | 42 | 29 | 29 | 37 | 42 | 29 | 39 | 34 |
| Control .. .. .. .. | 13 | 35 | 13 | 41 | 13 | 20 | 14 | 40 | 27 | 20 | 24 |

The experimental values represent all conditions taken together.

(3) There was considerable variation between different subjects. In the first session the interquartile range was 22·2 per cent. of the median, and in the second session 19·2 per cent. However there was a positive rank order correlation of 0·82 between the first and second sessions for each subject.

The finding of increased contrast with peripheral viewing was not confirmed by another experiment (Burgh and Grindley, 1962). In this case no difference in contrast was found. It must, however, be mentioned that in this experiment the patches were thrown onto the periphery by magnifying the entire display, so that the patches were separated by more than 1° of visual angle, and were also larger than in the condition of foveal viewing. Thus the conditions were not strictly comparable with those of the present experiment.

TABLE III

DEGREE OF CONTRAST, FIRST AND SECOND SESSIONS

MEAN OF ALL CONDITIONS

| *Subject* | 2 | 1 | 8 | 9 | 6 | 10 | 3 | 4 | 7 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|
| First session | 2·2 | 2·6 | 3·3 | 3·4 | 3·4 | 3·8 | 3·8 | 4·1 | 4·1 | 4·2 |
| Second session | 2·6 | 3·5 | 3·1 | 3·3 | 3·9 | 3·7 | 4·0 | 3·6 | 4·9 | 4·6 |

Rank-order correlation coefficient = 0·82.

It was however pointed out in discussion of the present experiment that the increase in contrast with peripheral viewing could be a result of the blurring involved in this condition. It has been reported (Tennant, 1929) that laying a sheet of tissue paper over a conventional contrast display enhances the apparent difference between the patches. Thus the loss of acuity involved in viewing the display peripherally might have a similar effect. Experiment II was therefore designed to test the effect of blurring upon simultaneous contrast.

## EXPERIMENT II

### *Display and Apparatus*

The display used was similar to that in experiment I, except that circular test and match fields were used instead of square ones.

The display was produced by two 2 in. × 2 in. slide projectors. One projected the test field and the two surrounds by means of a slide made from photographic film of suitable densities, while the other, which was fitted with adjustable crossed polaroids and a simple slide of sheet aluminium with a small hole, projected the match field onto the darker half of the ground. The diameter of each patch subtended 1° at the subject's eye and the patches were separated by 1°. A pencil mark on the screen immediately between the two patches was used as a fixation mark. The display was blurred by throwing the projectors out of focus. A piece of frosted glass was placed behind the slide in order to avoid casting an image of the projector lamp filament. The luminance gradient of the blur produced was scanned by a photometer and its shape is shown in Figure 3.

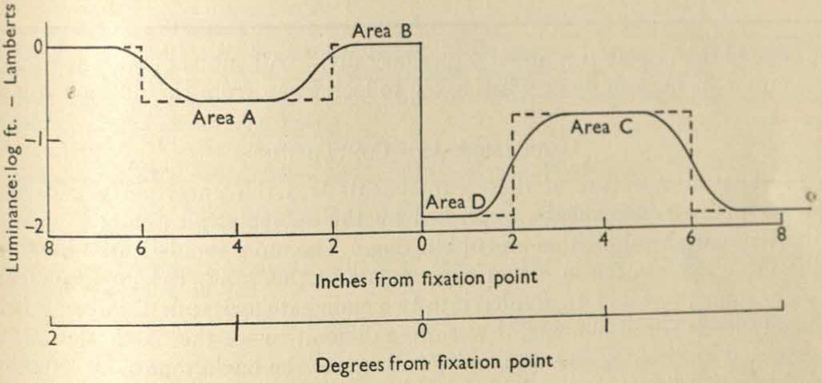


FIGURE 3

Photometric scan of display, showing changes of luminance in horizontal direction. Solid line, blurred condition. Broken line, sharp condition.

The luminances were as follows:

White surround (B) 0·8 ft. L.
Black surround (D) 0·047 ft. L.
Test field (A) 0·56 ft. L.
Match field (C) 0·04 to 0·18 ft. L.

The luminances were measured by means of an Ilford S.E.I. Photometer.

Six subjects were used, drawn from research workers of the Psychological Laboratory, Cambridge. Each made 20 matches under each condition. They were instructed to adjust the polaroid controlling the luminance of the match field until both patches appeared equally bright (in the blurred condition they were to match the centres of the patches) and to do this in 10 sec. The experimenter counted from a stop watch "5," "8, 9, 10." The polaroid was adjusted by means of a wheel in order to avoid kinaesthetic cues.

### *Results*

The results are shown in Table IV. All subjects showed more contrast in the blurred condition and these differences are significant by the Wilcoxon test ($p < 0{\cdot}001$). The scatter under sharp and blurred conditions is not however significantly different. This is somewhat surprising, as subjects expressed considerably less confidence in their matches when the display was blurred.

It is interesting that several subjects, who had not taken part in the first experiment, remarked spontaneously that contrast appeared to increase when they looked

TABLE IV

| | *Degree of contrast* | | | |
|---|---|---|---|---|
| | *Median* | | *Interquartile range as % of median* | |
| *Subject* | *Sharp* | *Blurred* | *Sharp* | *Blurred* |
| GCG | 9·2 | 11·4 | 16·3 | 8·7 |
| SA | 3·0 | 5·4 | 10·0 | 31·5 |
| CS | 6·3 | 9·4 | 44·5 | 34·0 |
| PW | 7·8 | 11·4 | 50·0 | 14·0 |
| SH | 4·6 | 7·6 | 8·7 | 11·8 |
| DS | 4·0 | 5·5 | 50·0 | 32·7 |
| Average | 5·8 | 8·4 | 29·9 | 22·1 |

at the edge of the display (i.e. about 6° peripherally). All subjects but one remarked on this either spontaneously or when asked to look away from the fixation point.

## DISCUSSION AND CONCLUSION

A most striking feature of these experiments was the uncertainty and lack of confidence in their judgements, expressed by the subjects, particularly under the conditions of peripheral fixation and of blurring. The more sophisticated the subject and the more experienced in visual experiments of this kind, the less confident he was. The difficulty could be divided into two main categories: first, in conditions of peripheral viewing or of blurring, it was often difficult to see the patches at all; they were reported as "waxing and waning," "fading into the background" or "appearing and disappearing." The second kind of difficulty consisted in a more or less ill-defined qualitative difference between the two patches, so that they never appeared the same. The patch on the black surround was reported as appearing "self-luminous," while the other, on the white surround, appeared as a patch of grey paper. Although at the two ends of the scale the two patches could confidently be reported as different in brightness, there was frequently no point in the middle where the patches were reported as being the same, but rather an area within which they were somehow different, although it was not possible to say in which direction the difference lay.

This difficulty suggests that there may be a deficiency in the concept of "brightness" as it is generally used in psychophysics. This is not to say that the trouble is "purely verbal" in the sense that re-defining the terms would eliminate the difficulty, but rather that the existence of the word "brightness" has led us to assume a simple sensory dimension analogous to the concept of luminance, whereas in fact the visual system may be making distinctions which are not reflected in our language.

It is evident that the nature of these distinctions should be established empirically, and the language of psychophysics, at least, be adapted in order to accommodate them. Until this is done the reliability of communication between experimenter and subject will remain doubtful.

Bearing in mind these fundamental objections to the concepts at present in use, it is not possible to draw any definite conclusions concerning the relation of contrast, blurring and peripheral fixation. There are three possibilities:

(1) That the increase of contrast in blurring and in peripheral fixation is due to a different, presumably retinal, mechanism in each case.

(2) That the increase found in peripheral fixation is due only to the loss of acuity involved, the basic connection being between blurring and increase of contrast.

(3) That the increase found in peripheral fixation and that found in blurring are both independent results of a general, possibly more central, mechanism controlling both contrast and "constancy"—another term which may require considerable re-thinking.

The selection of one of these alternatives would be guided by both physiological and psychophysical evidence; at present however the evidence is insufficient to allow such a selection to be made.

I wish to thank Mr. G. C. Grindley for frequent advice and discussion. I am also indebted to the Medical Research Council, under whose grant this research was carried out.

## References

Asher, H. (1950). Contrast in eye and brain. *Brit. J. Psychol.*, **40**, 187–94.

Barlow, H. B., FitzHugh, R., and Kuffler, S. W. (1957). Change of organization in receptive fields of the cat's retina during dark adaptation. *J. Physiol.*, **137**, 338–54.

Burgh, P. and Grindley, G. C. (1962). Size of test patch and simultaneous contrast. *Quart. J. exp. Psychol.*, **14**, 89–93.

Siegel, S. (1956). *Nonparametric Statistics*. London.

Tennant, J. (1929). The psychological factor in colour contrast. *Brit. J. Psychol.*, **20**, 1–26.

Wiesel, T. N. (1960). Receptive fields of ganglion cells in the cat's retina. *J. Physiol.*, **153**, 583–94.

Wilson, K. V. (1956). A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, **53**, 96–101.

*Manuscript received 8th March,* 1963.

# SHORTER ARTICLES AND NOTES

## NOISE AND THE SHIFTING OF ATTENTION

BY

W. M. S. SAMUEL

*From the Department of Psychology, Queen's College, Dundee*

An experiment was carried out to test the hypothesis that performance on a task which involves a shift of attention between spatially separated sources would be lower under noise conditions than one which involved no such shift. The results indicated that the hypothesis was not proved; rather they showed that the subject's performance on the shift of attention task was better under noise than under quiet. An attempt is made to put forward possible reasons for this occurrence.

### INTRODUCTION

A survey of theory and experimental data resulting from the study of noise and its effects, revealed that whilst the effects of noise on performance were not as widespread as might be imagined, certain types of task did reveal a deleterious effect. Broadbent (1953, 1954) showed that noise produced a decrement in the prolonged performance of vigilance and serial reaction tasks, and postulated that the significant feature of these tasks, which renders them prone to deteriorate under noise conditions, is that they are sensitive to momentary lapses of attention. When considering the work of Jerison (1954, 1957), whose task required the subject to press an appropriate key under three lights, flashing at different rates, when each had flashed a stated number of times, Broadbent (1958*b*) felt that, since under noise conditions the decrement increased as the interval between possible responses increased, this suggested that immediate memory could well be the mechanism which had broken down. To discover the validity of his proposition he conducted an experiment which placed a heavy load on subjects' immediate memory. The self-paced task was a difficult subtraction: a six-digit number appeared in a slot and was to be memorised before the subject moved the apparatus on to reveal a four-digit number in a second slot, the first number being no longer visible: the second number was to be subtracted from the first. Performance was significantly worse under 100 db. noise than under 70 db. noise.

Whilst Jerison (1957), in collaboration with Wing, found that there was a decrement under noise in the performance of a vigilance task consisting of three clocks, a further experiment (Jerison, 1957) using only one clock revealed no such decrement. In the light of this, Broadbent (1958*a*) observes: "This highly important result suggests that noise only affects tasks in which attention must be shifted regularly from one source to another: all the tasks showing effects are in fact of this type." He points out the distinction between internal and external shift of attention: by the former is meant the division of attention between immediate memory and the calculation of the response, and this is the aspect with which he has mainly concerned himself. The external shift of attention refers to the division of attention between sources of information. The spatial separation of these sources of information in noise experiments seemed to be a factor which required further experimental work, and it was hoped that the following experiment would delimit the factor sufficiently to reveal the effects of noise upon it.

### *Method and Apparatus*

Subjects faced a plain white screen, 104 × 102 cm., in which were cut two vertical slots, 30·5 cm. apart. The vertical length of the slots permitted adjustment for height so that the material would be directly in front of any subject. In the No Shift condition the subject sat in front of one slot, whilst in the Shift condition he sat in front of the mid-point between the two apertures: the chair was so placed as to have the subjects 1 m. away from the stimuli under both conditions. Two 60 watt adjustable reading lamps gave specific lighting to the displays.

The stimulus material consisted of digits typed on rolls of 35 mm. photographic printing paper. For the No Shift condition, two columns of 1,250 digits were typed one space apart on the same roll (Roll A). For the Shift condition, exactly the same digits were used but each column was typed on a separate roll of paper (Rolls B and C). The letters of the alphabet provided lead-in material on each roll.

The presentation device consisted of a solenoid activated by a microswitch attached to a cam, which provided the timing interval. From the solenoid extended a claw, which engaged in the holes along the edge of the paper as it was fed through guides. White noise was fed into calibrated earphones, worn by all subjects. A noise level of 110 db. was used in the Noise condition: 80 db. in the Quiet.

## RESULTS

| | | | | *Mean number per subject* | | |
|---|---|---|---|---|---|---|
| | | | *Error per cent.* | *Omissions* | *Wrong* | *Total responses* |
| Shift | Noise | F | 24 | 221 | 79 | 1,029 |
| | | M | 40 | 418 | 79 | 832 |
| | | F & M | 32 | 319 | 79 | 931 |
| | Quiet | F | 57 | 650 | 63 | 600 |
| | | M | 50 | 555 | 68 | 695 |
| | | F & M | 54 | 602 | 65 | 648 |
| No shift | Noise | F | 9 | 67 | 46 | 1,183 |
| | | M | 12 | 114 | 39 | 1,136 |
| | | F & M | 11 | 91 | 43 | 1,159 |
| | Quiet | F | 8 | 48 | 50 | 1,202 |
| | | M | 4 | 23 | 23 | 1,227 |
| | | F & M | 6 | 35 | 37 | 1,215 |

F—Female.
M—Male.
Wrong—Response made but calculation in error.
Omission—No response made to presentation.
Error per cent.—Wrong responses plus Omissions expressed as a percentage of total possible correct responses.

The subjects were volunteer students, randomly assigned to conditions. Ten subjects (one additional female was used in 1), five females and five males, served in each of the following four conditions:

(1) Shift with Noise.
(2) Shift with Quiet.
(3) No Shift with Noise.
(4) No Shift with Quiet.

In the Shift conditions, the subjects were required to add together the two digits as they appeared each second, in the two separated sources, on rolls B and C. In the No Shift conditions the subjects were required to add together the two digits as they appeared, one type space apart, in the single slot, on roll A only. The fronts of the displays were masked so as to allow only the two digits to be seen at any one time. The subjects spoke their answer to each addition, and this was checked by the experimenter against a master sheet. The duration of the experiment was 20 min. 50 sec., in which 1,250 presentations were made.

Three main points arise from these results:

(1) The hypothesis that there would be a lower performance under Noise in the Shift condition than under Quiet, is not proven.

### STATISTICAL ANALYSIS

| | *Significant difference* | | *No significant difference* |
|---|---|---|---|
| | 0·01 | 0·05 | |
| 1. Comparison between 21 Shift subjects and 20 No Shift subjects on basis of: | | | |
| Percentage Error Score | x | | |
| Omissions Score | x | | |
| Wrong Score | x | | |
| 2. (i) Comparison between conditions Shift-Noise and Shift-Quiet on basis of: | | | |
| Percentage Error Score | x | | |
| Omissions Score | x | | |
| Wrong Score | | | x |
| (ii) Comparison between conditions No Shift-Noise and No Shift-Quiet on basis of: | | | |
| Percentage Error Score | | | x |
| Omissions Score | | | x |
| Wrong Score | | | x |
| 3. (i) Comparison of scores between 11 females and 10 males in Shift and No Shift conditions under Noise, on basis of: | | | |
| Percentage Error Score | | x | |
| Omissions Score | | x | |
| Wrong Score | | | x |
| (ii) Comparison of scores between 11 females and 10 males in Shift and No Shift conditions under Quiet, on basis of: | | | |
| Percentage Error Score | | | x |
| Omissions Score | | | x |
| Wrong Score | | | x |

The Mann-Whitney Test (White's Modification) was used in all calculations.

Analysis of Total Responses follows the same pattern as Omissions Score, since the latter is equal to the Total Number of Possible Responses minus the Total Number of Responses.

(2) On the basis of Percentage Error Score and Omissions Score, the subjects performed significantly better under Noise on the Shift condition: but, for the Wrong Score, there is no significant difference between performance under Noise and that under Quiet. Even when the number wrong is rated in proportion to the number of respective responses, there is still no significant difference: this suggests that it is likely the process involved in this type of calculation is unaffected by noise. Consequently, on the basis of the foregoing, it would seem that either the perceptual or the effector mechanism is that which is affected by noise. Only further delimitation by experimental design, will provide a fuller answer to this problem.

(3) Differences between the performance of males and females are found notably under Noise, where females did significantly better than males, based on Percentage Error Score and Omissions Score: but no such difference was found using Wrong Score. There was also a significant difference, in favour of the males, based on Wrong Score in the performance of the No Shift task. The indication is that females were more highly activated by noise than the males, producing a greater output of responses. Since the number wrong appeared to be in the same proportion, on all conditions, to the total number of responses, the females achieved a better performance under Noise.

## DISCUSSION

The following features distinguish the present experiment from those which have shown a decrement under noise conditions:

(1) Simple work task.

(2) Externally paced at a high signal rate.

(3) Little load on immediate memory.

(4) No subject served in more than one condition.

(5) Shorter experimental period.

It was considered necessary to incorporate these features in the experimental design, to isolate as much as possible the effect of noise on the variable, "Shift of Attention." Bakan (1957) has shown that performance on a vigilance task is a function of the number of possible responses, and Jerison (1957) found that the order of subjection to noise and quiet influenced the results.

The following factors are put forward to explain why Noise influenced performance only in the Shift condition:

(1) Level of arousal.

(2) Compensatory effort.

(3) Deleterious effects of apparatus noise.

(4) Lack of auditory feed-back.

Due to the difficulty of the Shift condition, the arousal level may have been high initially, with noise seeking either to augment it or to arrest its decline. McBain (1961) suggested that increased variability in vigilance tasks may reduce the tendency for the subject's state of arousal to decline due to the monotonous experimental environment caused by high level white noise. Although the session length may be too short to show this effect, in the present experiment the more variable condition would appear to be that of Quiet, consisting of 80 db. noise, apparatus noise, caused by the click of the solenoid, and the subject's voice: the combined impact of these distractions may have produced an adverse effect. Woodhead (1959) found a deleterious effect of bursts of noise in decision making, but that the decrement occurred only immediately after the onset of noise. The solenoid noise may have had an impact similar to the bursts of noise. The noise level used in the Noise condition was sufficiently high to mask the apparatus noise and the sound of the subject's voice. Consequently there were no cues to presentation and the subject may not have been tempted to recheck his calculation in the Noise condition, with the result that the whole performance was more fluid.

Further experimental work to isolate the effects of noise on the variables which are probably in use in tasks which show a change in performance will have to be undertaken to clarify whether it is a single aspect of the psychological processes that is affected by noise, or whether it is a composite effect.

I wish to record my very grateful thanks to those who have helped to effect this work: especially to Dr. N. E. Loveless who supervised it; also to Dr. T. R. Lee, Mr. B. R. Singer, Mr. W. C. Crabb, and the subjects.

## REFERENCES

BAKAN, P. (1957). Extroversion-introversion and improvement in an auditory vigilance task. *M.R.C., APU Rep.*, 311/57.

BROADBENT, D. E. (1953). Noise, paced performance, and vigilance tasks. *Brit. J. Psychol.*, **44**, 295–303.

BROADBENT, D. E. (1954). Some effects of noise on visual performance *Quart. J. exp. Psychol.*, **6**, 1–5.

BROADBENT, D. E. (1958*a*). *Perception and Communication.* London.

BROADBENT, D. E. (1958*b*). Effect of noise on an "intellectual" task. *J. acoust. Soc. Amer.*, **30**, 824–7.

JERISON, H. J. (1954). Paced performance on a complex counting task under noise and fatigue conditions. *Amer. Psychologist*, **9**, 399.

JERISON, H. J. (1957). Performance on a simple vigilance task in noise and quiet. *J. acoust. Soc. Amer.*, **29**, 1163–5.

JERISON, H. J., and WING, S. (1957). Effects of noise and fatigue on a complex vigilance task. *Wright Air Development Center Tech. Rep.*, 57–14.

MCBAIN, W. N. (1961). Noise, the "arousal hypothesis," and monotonous work. *J. appl. Psychol.*, **45**, 309–17.

WOODHEAD, M. M. (1959). Effect of brief loud noise on decision making. *J. acoust. Soc. Amer.*, **31**, 1329–31.

*Manuscript received* 11*th January*, 1963.

# A FURTHER EXPERIMENT ON THE DISCRIMINATION OF OPEN AND CLOSED SHAPES BY RATS

BY

N. S. SUTHERLAND

*From the Institute of Experimental Psychology, Oxford University*

In a previous experiment on the visual discrimination of open and closed shapes by rats (Sutherland, Carr, and Mackintosh, 1962) it was found that (1) at the beginning of training the animals exhibited a strong preference for the more open shape; (2) shapes differing in their horizontal projections only were more discriminable than shapes differing in their vertical projections only. The present experiment is a replication of the earlier one using different shapes and different apparatus. The first result was fully confirmed and some confirmation was obtained for the second, although late in training the differences between groups trained with shapes differing on opposite projections were not statistically significant.

## INTRODUCTION

Figure 1 shows four pairs of shapes. The first member of each pair will be referred to as a closed shape (*C*), the second as an open (*O*): by a "closed" shape is meant one which is compact, by an "open" one which is spread out (for fuller discussion *v.* Sutherland, Carr, and Mackintosh, 1962). The first pair of shapes in each row will be called vertical shapes (*V*), the second pair horizontal (*H*). Sutherland, Carr, and Mackintosh (1962) trained rats to discriminate between the two pairs shown in the first row. They found that at first animals trained with open shapes positive made more correct choices than animals trained with closed shapes positive. Moreover, the groups with open positive performed above chance from the outset of training whereas the groups with closed positive at first performed below chance. This suggests that rats have a preference for open shapes. Of the groups with open positive, animals trained on the horizontal shapes performed better throughout training than animals trained with vertical shapes, whereas of the groups with the closed shape positive, animals trained on horizontal shapes at first performed worse than those trained on vertical shapes, later they performed better. It was concluded that the two horizontal shapes differ more from one another with respect to the classifying mechanisms in the rat's visual system than do the two vertical shapes. Such differences in the readiness with which different shapes are classified apart will be referred to as differences in "discriminability." It should be noted that differences in discriminability (as the word is used here) may be revealed not only by the ease with which an animal *learns* to attach differential responses to the members of different pairs of shapes, but also by the animal's tendency at the outset of training to give differential

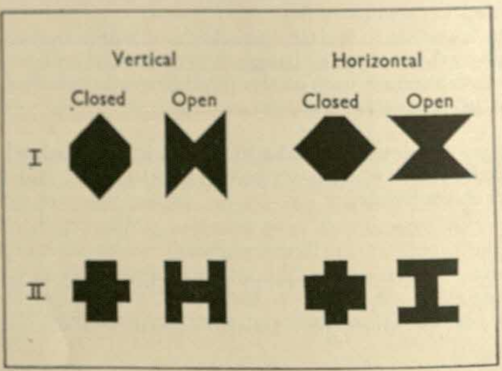


FIGURE 1

responses to shapes determined by its pre-existing preferences. The greater discriminability of the horizontal shapes is predicted by a theory suggested to account for shape recognition in the octopus (Sutherland, 1957, 1960), since the *V* shapes differ only in their vertical projections, the *H* only in their horizontal, and one assumption of the theory is that shapes differing in their horizontal projections only should be more readily discriminable than shapes differing in their vertical projections only.

Garner, Hake and Eriksen (1956) and Teuber (1960) have called attention to the importance of employing converging operations in the study of perception. The purpose of the present experiment was to replicate the earlier one using different shapes and a different training apparatus. The shapes used in the present experiment are shown in the second row of Figure 1. They were selected in such a way that one member of each pair was more open than the other, and one pair differs only in its vertical projections, the other only in its horizontal projections. The earlier experiment was performed in an apparatus in which a milk tube projected through the middle of each shape, and if animals ran to the positive shape they were able to drink from the milk tube. Since the rats' eyes are at the side of its head it is possible that the greater discriminability of the horizontal shapes was due to the rat seeing mainly the lateral parts of the shape whilst drinking: the lateral parts of the horizontal shapes differ more than the same parts of the vertical shapes. The apparatus used in the present experiment was a five door jumping stand: in this situation there is no reason to suppose that rats will attend more readily to the sides of a shape than to the top or bottom.

## METHOD

### *Subjects*

The subjects were 16 male hooded rats from the colony maintained at the Oxford University Institute of Experimental Psychology and they were about 6 months old at the start of the experiment.

### *Training procedure*

The apparatus was a conventional five hold jumping stand (cf. Fields, 1935). The shapes used were cut from ¼ in. white Perspex: they were of 24 sq. cm. area and could be attached to the front of doors in the jumping stand. Animals were run under 22·5 hr. hunger. They were first pretrained to jump to open doors, and then to a closed door bearing the positive shape alone. They were then given 300 training trials on which both positive and negative shapes were exposed. The two shapes were always exposed on adjoining doors, and the pair of doors used and the position (left or right) of the positive shape was varied randomly from trial to trial. Animals were always placed at the beginning of a trial either at the extreme left or the extreme right of the jumping platform: the position in which they were placed was randomized for each pair of doors used and for the position (left or right) of the positive shape. Correct jumps were rewarded with 20 sec. access to food. Ten trials were given a day and the intertrial interval was about 12 min. A modified correction procedure was used throughout, i.e. if animals jumped to the incorrect door they were replaced on the platform and continued to jump until they made a correct choice. In computing scores only first trial errors were counted.

### *Experimental design*

There were four groups each of four animals. (1) Group *VC* were trained with the vertical shapes (left hand pair row II of Fig. 1) with closed positive. (2) Group *VO* were trained with the vertical shapes with open positive. (3) Group *HC* were trained with the horizontal shapes (right hand pair of row II in Fig. 1) with closed positive. (4) Group *HO* were trained with the horizontal shapes with open positive.

## RESULTS

Learning curves for the different groups are shown in Figure 2. In the previous experiment, the number of correct choices made by the different groups early in training was in the order $HO > VO > VC > HC$. This order is reproduced exactly in the results of the present experiment over the first 12 days' training. Moreover, of the four comparisons which are important three were significant at better than the 0·02 level of confidence by a nonparametric test: thus every animal in Group *HO* scored better than any animal in Group *HC*, every animal in *VO* scored better than any in *VC*, and every animal in *VC* scored better than any in *HC*. The difference between *HO* and *VO* although in the expected direction, was not, however significant (t = 1·4, d.f. = 6).

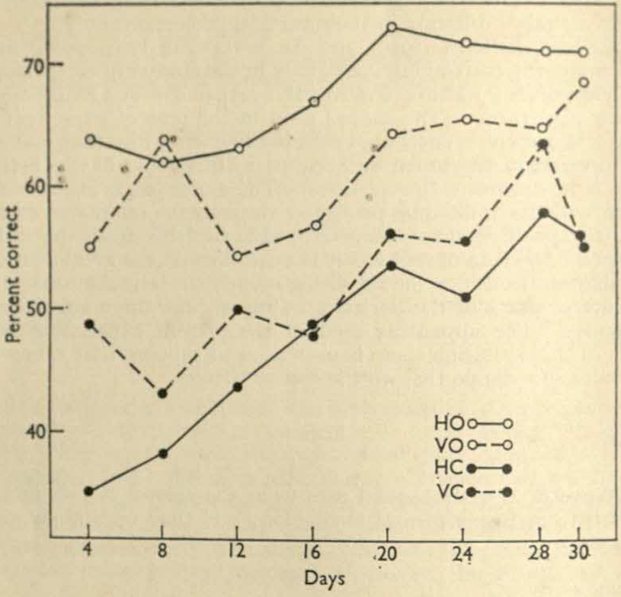


FIGURE 2

Later in training in the previous experiment the scores of the different groups were ordered as follows: *HO* = *HC* > *VO* = *VC*. In the present experiment *HO* continues to score better than either of the *V* groups, *HC* comes to equal *VC* but fails to surpass either *VC* or *VO*. Over days 13–30, there was no overlap between the scores of the two *H* groups nor between those of the two *V* groups: the differences between *HO* and *VO* and between *HC* and *VC* were not significant.

Thus the results early in training fully confirm those of the previous experiment and indicate that rats have a strong preference for open shapes. They also suggest that the horizontal shapes are more discriminable than the vertical ones since the preferences shown are much stronger for the *H* groups than for the *V* and the difference between the scores of *HC* and *VC* was significant. The results later in training are less clearcut. They again indicate strong preferences for open shapes, but the *H* groups failed to perform significantly better than the corresponding *V* groups. It seems likely that the reason why the results of later training are not completely in line with those of the previous experiment is that in the present apparatus learning was extremely slow. The reason for this is not known. Despite the lack of significant differences it should be noted that the results late in training confirm those of the earlier experiment in that *HO* continues to do better than *VO*, whereas *HC* does catch up with *VC*.

The agreement between the present results and those of the earlier experiment suggests that the previous results were not a function of the apparatus used and that they can be generalized to other pairs of open and closed shapes differing in either their vertical or horizontal projections. The present results fully corroborate that rats have an initial preference for open over closed shapes, and they strongly suggest that shapes differing in their horizontal projections are more readily discriminable than shapes differing in their vertical projections.

Since this experiment was performed, Barnes and Baron (1961) have reported an experiment which shows that mice also prefer more open (or more complex) shapes. Their animals learned to press a bar which exposed a shape, and they pressed more readily for a cross than for an outline square, and more readily for an outline square than for a filled-in circle.

This work forms part of a project on "Stimulus analysing mechanisms" and was supported by the American Office of Naval Research (Contract N62558-2453) and by the Nuffield Foundation.

## REFERENCES

BARNES, G. W., and BARON, A. (1961). Stimulus complexity and sensory reinforcement. *J. comp. physiol. Psychol.*, **56**, 466–9.

FIELDS, P. E. (1935). Studies in concept formation: II. A new multiple stimulus jumping apparatus for visual figure discrimination. *J. comp. Psychol.*, **20**, 183–203.

GARNER, W. A., HAKE, H. W., and ERIKSEN, C. W. (1956). Operationism and the concept of perception. *Psychol. Rev.*, **63**, 149–59.

SUTHERLAND, N. S. (1957). Visual discrimination of orientation and shape by *Octopus*. *Nature*, **179**, 11–3.

SUTHERLAND, N. S. (1960). Theories of shape discrimination in *Octopus*. *Nature*, **186**, 840–4.

SUTHERLAND, N. S., CARR, A. E., and MACKINTOSH, J. L. (1962). Visual discrimination of open and closed shapes by rats. I. Training. *Quart. J. exp. Psychol.*, **14**, 129–39.

TEUBER, H-L. (1960). Perception. In *Handbook of Psychology, Section 1: Neurophysiology, Vol. 3*, Chap. 55, pp. 1595–1660. Amer. Physiol. Soc.: Washington, D.C.

*Manuscript received 21st December, 1962.*

# VISUAL DISCRIMINATION OF SHAPE BY HUMANS

BY

J. BUTLER

*From the Psychological Laboratory, University of Exeter*

Reaction times to tachistoscopic exposure and a card-sorting task are the two methods by which visual discrimination of shape is studied. The results of experiments using both methods tally, i.e. the order of difficulty of the following shapes was: (1) (easiest) circle vs. triangle; (2) circle vs. square; (3) square vs. triangle; (4) (most difficult) triangle vs. diamond. These results are compared with those obtained by Sutherland, using octopus, and it is concluded that it should be possible to establish to what extent the system of visual analysis in humans differs from that postulated for both octopus and rat.

## INTRODUCTION

Although the discrimination of shape by sub-human animals is currently being studied (see, for example, Sutherland, 1960), so far as the author is aware, no comparable work has been done of late, using human subjects. The object of this paper, therefore, is to examine, by two different methods, visual discrimination of shape by humans. A reaction time experiment will be described and, after that, a card-sorting one. Both experiments are based on three assumptions. (i) By these methods, only one dimension is under analysis, i.e. that all shapes are translated into a finite amount of excitation. (ii) Shapes would thus be rangeable along a single continuum. (This assumption seems justified when one considers that mensuration demands the emphasis of one characteristic to the exclusion of others: thus, weighing ignores length, while the measurement of length is not concerned with weight.) The hypothesis was that, when there was a "large difference" between the two shapes, the response time (or sorting time) would be short, and when there was a "small difference" between the two shapes, the response time (or sorting time) would be long. This is concluded from the assumption (iii) that large degrees of difference are soon discerned by the subject, while small degrees of difference require comparison of the various characteristics of the shape, which is presumed to take more time. It should be noted that no verbal definition of what constitutes "difference" or "identity" is required here.

## EXPERIMENT I

*Method.* Pairs of shapes were displayed to human subjects tachistoscopically. The instrument (designed and constructed by Dr. A. Staniland) presented the shapes through superimposed red and green filters (which markedly reduced their luminosity) thus: the shapes were 36 cm. from the eye, subtending a visual angle of 1° 54′ approximately. A warning noise preceded the exposure of the first shape, the exposure lasting one-fifth of a sec. (the fractional exposure); there was an interval of 3 sec. before the second shape appeared. Exposure of the second shape lasted until the correct response was made. The subject was required to respond by pressing a key marked "S" for two same shapes or a key marked "D" for two different shapes. If the wrong key was pressed, the exposure continued until the experimenter himself ended it, and no time was taken for responses of this sort. The subject was not allowed further attempts to correct errors, save when the same pairs recurred in their randomized sequence. These errors, which were negligible in number, were discarded. Responses were timed to the nearest hundredth of a second on a Venner stop-clock, capable of recording ten-thousandths of a sec.

The shapes used (which were solid, not outline) were taken from Sutherland's work (1957) in order that close analogies could be drawn; and these shapes were scaled down (one-tenth of an inch representing 2 cm.) to fit standard Kodak 35 mm. slides and drawn in pencil on transparent stiff paper. Eight pairs of different shapes (e.g. ◇ vs. △ or ⊢ vs. ⊣) and eight pairs of same shapes (e.g. ◇ vs. ◇ or □ vs. □) were presented and the order of presentation for all 16 pairs was randomized. There is no evidence to suggest that which member of a pair was shown first had any effect on the results.

The subjects, 18 in number, were experimentally naïve female students of Occupational Therapy and their teachers, of between 18 and 28 years and, in all cases, a practice period of 20 trials was allowed. (The results given below have had this period abstracted.) This precaution was taken so that there could be no doubt that each subject knew what constituted a "same" decision and what a "different" one. Such procedure obviated the need for detailed instructions to subjects, who were merely asked to respond to differences of shape and orientation.

TABLE I

EXPERIMENT I. 18 SUBJECTS. MEAN REACTION TIMES: INDIVIDUAL AND TOTAL (IN SEC.)

| INDIVIDUAL | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 1·305 | 1·450 | 1·180 | 1·345 | 1·285 | 1·325 | 1·160 | 1·225 |
| 2 | 1·210 | 1·435 | 1·170 | 1·025 | 1·050 | 1·255 | 1·190 | 0·960 |
| 3 | 1·730 | 1·420 | 1·140 | 1·385 | 1·415 | 1·165 | 1·220 | 1·125 |
| 4 | 2·650 | 1·350 | 2·225 | 1·230 | 1·320 | 1·220 | 1·220 | 1·125 |
| 5 | 1·240 | 1·360 | 1·170 | 1·230 | 1·310 | 1·050 | 1·040 | 1·160 |
| 6 | 1·305 | 1·535 | 1·105 | 1·380 | 1·120 | 1·530 | 1·355 | 1·295 |
| 7 | 1·245 | 1·175 | 1·500 | 1·100 | 1·105 | 1·110 | 1·030 | 1·055 |
| 8 | 1·535 | 2·000 | 1·605 | 2·300 | 1·760 | 1·320 | 0·900 | 0·935 |
| 9 | 1·520 | 1·050 | 1·160 | 0·960 | 0·890 | 0·925 | 0·830 | 0·835 |
| 10 | 1·400 | 1·420 | 1·180 | 1·100 | 1·050 | 1·360 | 1·065 | 0·945 |
| 11 | 1·540 | 1·420 | 1·250 | 1·230 | 1·050 | 1·120 | 1·040 | 1·035 |
| 12 | 1·340 | 1·780 | 1·220 | 1·310 | 1·410 | 1·085 | 1·090 | 1·140 |
| 13 | 1·085 | 1·360 | 0·785 | 0·825 | 0·980 | 0·900 | 0·975 | 0·675 |
| 14 | 0·945 | 0·760 | 0·755 | 0·755 | 0·740 | 0·695 | 0·750 | 0·700 |
| 15 | 0·715 | 0·800 | 0·600 | 0·650 | 0·790 | 0·800 | 0·800 | 0·630 |
| 16 | 1·360 | 1·260 | 1·345 | 1·250 | 1·350 | 1·320 | 0·970 | 1·140 |
| 17 | 0·970 | 1·250 | 0·915 | 0·860 | 1·130 | 1·000 | 0·880 | 0·770 |
| 18 | 1·690 | 1·110 | 1·100 | 0·970 | 0·990 | 0·990 | 1·170 | 1·040 |
| TOTAL | 1·377 | 1·330 | 1·189 | 1·161 | 1·152 | 1·131 | 1·043 | 0·988 |
| SHAPES | △ ◇ | ⊢ ⊣ | — / | \| / | T ⊥ | □ △ | ○ □ | ○ △ |

The difference between two adjacent groups of the following pairs of shapes is significant at better than the level of confidence indicated (Wilcoxon matched pairs test—two tailed):

(1) diamond vs. triangle (0·01); (2) square vs. triangle (0·05); (3) circle vs. square; (4) circle vs. triangle; (5) ⊢ vs. ⊣ (0·01); (6) T vs. ⊥.

N.B.—The average reaction times to "same" shapes was as follows: diamond vs. diamond: 1·257 sec.; square vs. square: 1·248 sec.; ⊢ vs. ⊢: 1·213 sec.; circle vs. circle: 1·179 sec.s | vs. |: 1·175 sec.; / vs. / 1·143 sec.; T vs. T: 1·094 sec.; — vs. —: 1·093 sec. The purpose of the inclusion of "same" shapes in the experimental design was to provide a situation where there were equal numbers of "same" and "different" stimuli.

## DISCUSSION

### *Squares, circles and triangles*

It will be observed that these results differ in detail from those obtained by Sutherland, using octopus; based on percentage correct responses (cf. Sutherland, 1957) the shapes presented an order of difficulty of discrimination as follows: diamond vs. triangle and circle vs. triangle (both 67 per cent.); circle vs. square (69 per cent.); circle vs. diamond (74 per cent.); square vs. triangle (85 per cent.). Elsewhere, Sutherland (1959) reports that circle vs. square, circle vs. diamond and circle vs. triangle were equally discriminable for the octopus. Thus, results with both octopus and humans show that diamond and triangle are much less readily discriminable than square and triangle. Now, Sutherland has brought forward a number of suggestions as to how visual stimuli may be analysed by octopus. The most fundamental of these is that the octopus discriminates shape by analysing the horizontal and vertical extents of figures. A neurological mechanism is proposed "which counts horizontal extents of a shape at each point on the vertical axis, and vertical extents of a shape at each point on the horizontal axis" (1960). Such a proposal is in line with a large number of Sutherland's experimental findings. Further, it would appear from the results that the horizontal projection is more accurately analysed than the vertical. However, results with human subjects cannot be explained on the basis of the vertical and horizontal projections of the shapes concerned.

*"T" figures and rectangles*

The results with human subjects show that up–down mirror image shapes are more readily discriminable than left–right mirror image shapes. Sutherland has found this to be the case with octopus.

There is not any significant difference between the time taken to respond to — vs. / and | vs. /. This agrees to some extent with Sutherland's results. (Cf. 1960: "Animals trained to discriminate between a horizontal and oblique treat a vertical rectangle as an oblique, and vice versa.") It is worth noting in this connection that the mean response time (18 subjects) for two "same" oblique shapes was 1·143 sec., while that for the discrimination | vs. / was 1·161 sec. This difference is minute.

It was thought that the value of these results (with human subjects) depended on the exclusion of any interpolation that might be affecting them. A number of objections to the method may be made. To give two instances:

(i) The temporal arrangement of displaying shapes, i.e. fractional exposure, lasting one-fifth of a sec., 3 sec. pause, then the other exposure—this arrangement is an arbitrary one. Any part or parts of this sequence may interfere with the response time.

(ii) Some artefact, or cue (other than the shapes), might be prompting the responses, unknown to the experimenter.

Thus, a card-sorting experiment was devised to test the method of Experiment I.

## EXPERIMENT II

*Method.* The cards used were blank playing cards of standard size, on the face of which black shapes (solid, not outline) were glued. The shapes (square, diamond, triangle, circle) were equal in area (25 sq. cm.) There were 20 subjects, divided into two groups: Group A (11 subjects) and Group B (9 subjects). Group A worked with 15 cards of each "suit," together with 15 blank cards (i.e. a pack of 45 cards with, say, 15 squares, 15 circles and 15 blanks to sort into three piles); Group B worked without any blank cards (i.e. using a pack of 30 cards, say, 15 diamonds and 15 triangles, to sort into two piles). Group A was given the blanks to test what effect a third category would have on the results. It was found that, when the number of subjects is small, blank cards hardly affect the differences or relative sorting times. The categories for sorting were: (i) circle from triangle; (ii) circle from square; (iii) square from triangle; (iv) diamond from triangle. There were five trials per category and the categories were presented to the subjects in random order. Inter-trial shuffling lasted between 30 sec. and 1 min., but no standard method of shuffling was adopted. The subjects (again, students and teachers of occupational therapy; but no person did both experiments) were instructed to sort the cards into "suits" as quickly as they could and were timed by a manually operated stop-watch.

## RESULTS AND DISCUSSION

The card-sorting results produce the same temporal order as those of the reaction time experiment: in *both* experiments the speed of discrimination was: (1) circle vs. triangle; (2) circle vs. square; (3) square vs. triangle; (4) triangle vs. diamond; and thus the tachistoscopic method seems to be vindicated.

Several comments of a theoretical nature remain to be made.

(*a*) No "explanation" of the results, in terms of postulated neural arrangements, is offered at this stage. Primarily this is because the experimental facts are few and the devisable hypotheses that would cover them are many. This work is designed to facilitate some inter-species comparisons and it is thought that data of a comparative nature are of more importance than further postulates. What is worthy of note, therefore, is that, although the results with square, triangle, circle and diamond cannot be subsumed under

Sutherland's suggested analysis of vertical and horizontal extents, *this is equally the case with the octopus results.* Anyone who cares to draw the projections for these shapes (in the same way that Sutherland does, i.e. in 1960) will see that both vertical and horizontal projections of triangle, circle, and diamond are very similar whereas those for the square are peculiar to that figure. This does not detract much from Sutherland's theorizing, since he has merely suggested that analysis into projections is only a first and approximate step, one that was prompted by earlier findings; that he is aware of the difficulties involved in the next step is shown by his published work.

TABLE II

EXPERIMENT II. 20 SUBJECTS. MEAN SORTING TIMES: INDIVIDUAL AND TOTAL (IN SEC.)

| *Sorting categories* | *Triangle from diamond* | *Square from triangle* | *Circle from square* | *Circle from triangle* |
|---|---|---|---|---|
| INDIVIDUAL | | | | |
| Group A | | | | |
| 1 | 43·32 | 40·28 | 38·84 | 38·60 |
| 2 | 39·88 | 37·08 | 36·16 | 36·60 |
| 3 | 49·00 | 43·68 | 39·68 | 42·16 |
| 4 | 42·24 | 41·32 | 40·76 | 40·08 |
| 5 | 42·44 | 39·64 | 40·44 | 39·24 |
| 6 | 46·28 | 45·20 | 44·00 | 44·64 |
| 7 | 38·56 | 37·68 | 35·04 | 36·80 |
| 8 | 37·12 | 34·20 | 33·12 | 32·96 |
| 9 | 45·48 | 43·04 | 43·04 | 40·30 |
| 10 | 41·48 | 39·64 | 38·64 | 37·60 |
| 11 | 44·88 (0·01) | 42·24 (0·01) | 41·64 | 41·52 |
| Group B | | | | |
| 1 | 32·56 | 30·72 | 30·40 | 29·84 |
| 2 | 35·48 | 34·56 | 32·36 | 33·36 |
| 3 | 26·76 | 25·80 | 24·56 | 24·80 |
| 4 | 27·20 | 25·80 | 25·00 | 24·10 |
| 5 | 27·72 | 26·20 | 26·04 | 24·52 |
| 6 | 28·32 | 27·44 | 27·20 | 26·48 |
| 7 | 34·30 | 31·50 | 30·00 | 29·40 |
| 8 | 27·88 | 26·68 | 25·96 | 25·28 |
| 9 | 31·28 (0·01) | 29·52 (0·01) | 26·96 | 26·08 |
| TOTALS | | | | |
| Group A | 42·79 | 40·36 | 39·21 | 39·13 |
| Group B | 30·16 | 28·69 | 27·60 | 27·09 |

The difference in performance between two adjacent categories is better than the level of confidence indicated (Wilcoxon).

(*b*) The possibility has been suggested that the discriminability of one pair of shapes may depend on what other shapes are included in the experimental ensemble. Should this be happening, it would be difficult to detect in the experiments described; but, in future, by comparing the results of different experimental sets of shapes with a common factor, the criticism could be answered.

(*c*) By the methods outlined above it will be possible to test whether Sutherland's basic hypotheses are relevant to human vision. *If* one postulates that humans analyse shape in terms of vertical and horizontal projections, one is committed to the view that horizontal extents are more accurately analysed than vertical extents, because of the finding that up–down mirror images are more readily discriminable than left–right mirror images. (Up–down mirror images have a common vertical projection, but different horizontal ones; left–right mirror images have a common horizontal projection, but different vertical ones.) Further, it may well be the case that, consequent upon adopting the hypothesis that in human vision the basic analysis is in terms of horizontal and vertical projections, the same difficulties (i.e. of "further analysis") will come to light. Should, however, this postulate be later found to be experimentally acceptable, essentially the same system would have been suggested for octopus, rat (Sutherland and Carr, 1962) and man. The exploration of possibilities of this kind is the main purpose of the methods described here.

The author wishes to thank Dr. J. Szafran, Dr. A. Staniland, and Dr. R. Lynn for their supervision, criticisms and advice with regard to this paper.

## REFERENCES

SUTHERLAND, N. S. (1957). Visual discrimination of orientation and shape by *Octopus*. *Nature*, **179**, 11–3.

SUTHERLAND, N. S. (1959). Visual discrimination of shape by *Octopus:* Circles and squares, and circles and triangles. *Quart. J. exp. Psychol.*, **11**, 24–32.

SUTHERLAND, N. S. (1960). Theories of shape discrimination in *Octopus*. *Nature*, **186**, 840–4.

SUTHERLAND, N. S., and CARR, A. E. (1962). Visual discrimination of open and closed shapes by rats. II. Transfer tests. *Quart. J. exp. Psychol.*, **14**, 140–56.

*Manuscript received 23rd October, 1962.*

# APPARENT DISTANCE IN A HORIZONTAL PLANE WITH TACTILE-KINESTHETIC STIMULI

BY

ROBERT S. DAVIDON and MEI-FANG HSIEH CHENG

*From Bryn Mawr College*

In the study of active tactile-kinesthetic space perception apparent distance has been found to vary as a function of the direction of the line segment in a horizontal plane. The present data indicate that the error is not consistently related to the same frame of reference as the visual illusion. Rather, with movement of the extended arm to determine the relative distances between pairs of points, radial distances are overestimated in relation to tangential ones, whether parallel or perpendicular to the medial plane. Interpretations are in terms of kinesthetic stimulus patterns and the structure of perceptual representation.

## INTRODUCTION

A stimulus pattern which elicits a predictable constant error in judgements of relative length can be labelled an illusion. The "error" can be of greater psychological significance, however, if it can be identified as an instance of a general psychophysical relation within a perceived field.

Reid (1954) demonstrated and named an "illusion of movement." The blindfolded subject moved a stylus a standard distance to the corner of a horizontal square frame, and then reproduced an apparently equivalent distance on the adjacent side. "Vertical" (medial) distances were overestimated in relation to horizontal (frontal) ones. Reid accepted this constant error in distance judgements with arm movements to be the complement of the visual horizontal-vertical illusion and considered the possibility that the former might account for the origin of the latter.

In the visual illusion the reference for vertical is the medial plane of the head, not the gravitational vertical. Künnapas (1958) has demonstrated that tilting the subject's head 85° from the vertical leads to a reversal of relative length of segments of the same L-shaped figure. He has proposed that the oval shape of the visual field may account for the constant errors in horizontal-vertical judgements. Modifying the shape of the field has had an effect, and virtually eliminating the visual boundaries by having the judgements made in a dark room has reduced the illusion (Künnapas, 1955, 1957).

There is reason to question whether the constant error demonstrated in the illusion of movement, without vision, is consistently related to the medial plane of the head or to the boundaries of the visual field. With active tactile-kinesthetic stimulation what parameters of the field are relevant to the constant error in apparent distance?

In Reid's experiment with movement of the styli in a horizontal plane the orientation of the distances to be judged, parallel or perpendicular to the medial plane, was confounded with type of arm movement. "Vertical" distances were indicated by a radial movement, involving some flexion or extension (the term vertical is in quotation marks for the segment was not vertical in the usual sense). Horizontal distances were denoted by a movement tangential to an arc with the observer at the centre, involving some adduction or abduction.

In the present experiment orientation to the medial plane is varied independently of the type of movement. A distance between a pair of points to the front is compared to that between two at the side. The same type of movement is used for determining a radial distance in front, parallel to the medial plane ("vertical"), as is used for a radial distance at the side perpendicular to it (horizontal). Similarly, a tangential movement may be either perpendicular to the medial plane at the front (horizontal) or parallel to it at the side ("vertical"). Note Condition VIII in Figure 1.

Two contrary hypotheses are tested:

(1) A "vertical" distance will be overestimated in relation to a horizontal one, if the function is the complement of the visual horizontal-vertical illusion. When parallel lengths (both "vertical" or both horizontal) are compared there will be no significant error consistent from subject to subject.

(2) A radial distance will be overestimated in relation to a tangential one, if the type of arm movement is crucial and Reid's findings can be generalized. When two perpendicular radial or perpendicular tangential lengths are compared there will be no significant error consistent from subject to subject.

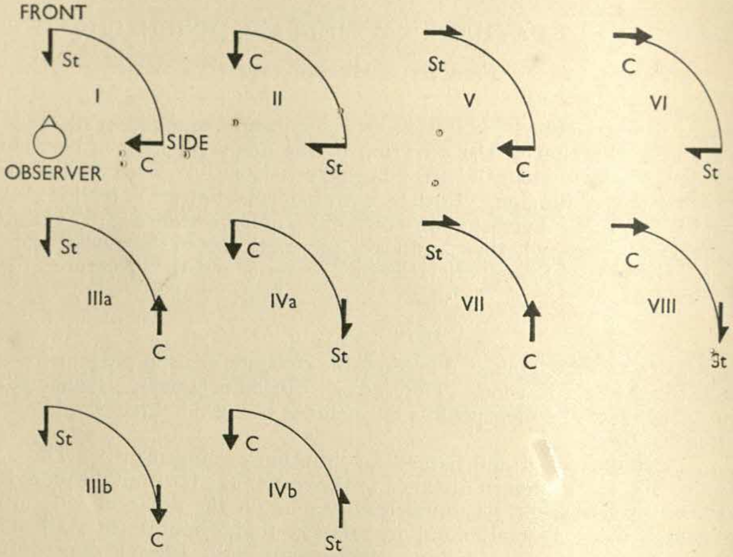


FIGURE 1

Conditions of presentation of the standard (St) and comparison (C) stimuli. For the comparison the arrow indicates the direction of movement of the variable stylus. For the standard the arrow indicates the order in which the two fixed styli were touched.

A significant positive constant error in Conditions I and VIII, a negative error in Conditions II and VII, and no significant constant error for Conditions III, IV, V and VI would confirm the former hypothesis; a significant positive effect for Conditions III and V, a negative error in Conditions IV and VI, and no significant constant error for Conditions I, II, VII and VIII would confirm the latter (see Fig. 1).

With either of the above hypotheses the constant errors in distance judgements can be interpreted as a set of functional relationships in a horizontal plane which define a general characteristic of the tactile-kinesthetic spatial schema.

## METHOD

*Subjects.* Ten women, graduate and undergraduate students at Bryn Mawr College, were paid for participation.

*Apparatus.* On a large table-top before the seated subject were two pairs of vertical styli, one pair to the front and one to the right at arm's length. Each stylus, a pointed ¼ in. dowel 2 in. tall, was mounted on a rectangular base. One pair was secured to the table during each group of experimental trials to provide a standard distance of 10 cm. between points. To provide the variable distance, one stylus of the other pair was fixed and one was free to slide between parallel guides.

*Design and procedure.* The method of adjustment was used for each judgement of equivalent distances between the pair of styli to the front and the pair at the side. There were 10 stimulus conditions with each subject making repeated judgements under each condition. The conditions differed in the orientation of the pairs to the medial plane, either parallel or perpendicular, and in the position of the standard at the front or side—as indicated in Figure 1. Conditions IIIb and IVb provided supplemental control, to determine possible effects of direction of manipulation of the variable and standard styli.

Within each experimental session (day) the sequence of stimulus conditions was determined by a random method, and two judgements were made under each condition. Data were obtained on five successive days for each of the conditions.

For each subject there was one preliminary or practice session with four trials in which a 12 cm. standard was used and the orientation of the variable was 45° rather than 0° or 90°.

The blindfolded subject was seated and her position and chair-height adjusted so that both the far stylus to the front and that at the side could be readily reached by the extended right arm at shoulder level. The experimenter guided the subject's hand so that the tips of the thumb and fingers could first surround the top of one stylus of the standard pair and then the other. He said, "This is the standard distance." After a second indication of the standard distance the subject's hand was guided to the styli of the variable pair in turn, and she was told which stylus was fixed and which movable. She was asked to increase and adjust the distance between the styli of the variable pair until they were equally far apart as the standard pair. The standard distance was always presented first. On a given observation the two styli of a pair could be successively touched as often as desired, but after an adjustment of the variable had been made, the subject could return to the standard no more than twice. No time limit was imposed. When necessary the experimenter guided the subject's hand to locate the standard and variable pairs.

## RESULTS

The experimental treatment, variation of the orientation and position of pairs of points to be judged, produced highly significant effects. In a treatments by subjects analysis of variance: $F = 24{\cdot}16$; $d.f. = 7, 63$; $p < 0{\cdot}001$. When the distances compared were parallel to each other, either both parallel to the median plane or both perpendicular to it (Conditions IIIa, IIIb, IVa, IVb, V and VI), the radial distance was consistently overestimated in relation to the tangential one (see Table I). In Conditions III*a*, III*b* and VI with a radial standard 279 of 300 judgements were overestimates; while with a tangential standard, Conditions IV*a*, IV*b*, and V, 300 of 300 judgements were underestimates.

TABLE I

VARIABLE DISTANCE JUDGED EQUIVALENT TO 10 CM. STANDARD (N = 10)

| *Orientation of standard* | | *Relation of standard and comparison stimuli* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *"Vertical"-horizontal* | | | *Both "vertical"* | | | *Both horizontal* | | |
| | | *Cond.* | *Mean* | *s. d.* | *Cond.* | *Mean* | *s. d.* | *Cond.* | *Mean* | *s. d.* |
| Radial | Front | I | 9·98 | 0·67 | III*a* | 11·50 | 1·01 | VI | 11·93 | 1·32 |
| | Side | II | 9·52 | 0·76 | III*b* | 11·85 | 1·28 | | | |
| Tangential | Front | VII | 8·66 | 0·71 | IV*a* | 8·14 | 0·78 | V | 7·62 | 0·92 |
| | Side | VIII | 10·11 | 0·76 | IV*b* | 7·92 | 1·00 | | | |

Comparison of individual means by Duncan's Multiple Range Test indicated that the differences between Conditions III*a* and IV*a*, III*b* and IV*b*, and between V and VI were significant ($p < 0{\cdot}01$). The second hypothesis is consistently confirmed, while the first is not supported.

Differences between Conditions III*a* and VI and between IV*a* and V were not significant, indicating that whether the standard was at the side or front produced no reliable effect.

When the standard and comparison were perpendicular to each other and both radial (Conditions I and II) there was no evidence of a constant error, judgements were often reversed on different days, and the two condition means did not differ significantly. No horizontal-vertical illusion was evident. With the pairs of styli perpendicular to each other and both tangential there was a tendency to underestimate the standard when it was horizontal (Condition VII), but no significant overestimation when it was vertical (Condition VIII). With no evidence of a constant error under three of the four critical conditions, the first hypothesis is rejected.

Finally, as indicated by a comparison of Condition IIIb with Condition IIIa, and of Condition IVb with Condition IVa, the direction in which the variable was manipulated or the standard touched had no significant effects.

## Discussion

The data confirm that with tactile-kinesthetic, as with visual, stimuli there are constant errors in relative distance comparisons which are a function of orientation within the field. To account for the consistent patterns of judgements, however, different reference systems are required for the two modalities. What Reid termed the "illusion of movement" is not simply the complement of the horizontal-vertical illusion. Without vision, whether the line segment is parallel or perpendicular to the medial plane is not crucial for predicting a constant error in relative length. Rather, the error is associated with radial and tangential directions within a polar schema.

Although one might extensively measure psychophysical functions and speculate no further, to order future investigations of space perception, both without and with vision, it will be tentatively assumed that the observed constant errors reflect differences between the array of stimuli *on the one hand*, and the structure of their representation within the organism on the other. The term representation may refer either to the extensional organization of afferent activity at a given level of the central nervous system or to the "phenomenal world."

As a first order explanation of the data of this experiment it might be proposed that actions of different groups of muscles resulting in equal displacement of the finger tips may be *represented* unequally. Lateral movements of the extended arm do involve distinctly different patterns of muscle activation and motion at the joints than radial movements which include flexion. Further studies are required of normal perception of proximal distances, under varied conditions of kinesthetic and tactile stimulation. No systematically collected body of quantitative data exists, although there are many reports of size and distance distortions associated with central nervous system pathology.

Were the demonstrated relation between orientation and apparent distance to be maintained under widely different conditions of kinesthetic and haptic stimulation, including judgements of tactile stimuli with the observer immobile, higher order properties of the representation would be indicated. As in theories of visual space perception, relations would then be discussed in the language of field effects.

Is there a unitary representation at some level of brain function to provide a common structure for all space perception regardless of modality? Kelvin's (1954) first comparison of tactual and visual judgements of length gave similar points of subjective equality, which he interpreted as reflecting a single schema. In a later study (Kelvin and Mulik, 1958), however, when he varied the relation of the midpoint of the comparison series to the standard the effects were not the same for intermodal as for ipsi-modal comparisons and he rejected his earlier explanation. Sandström (1951) has shown repeatedly that under conditions of restricted vision there is a consistent difference in the locus of a point in visual and kinesthetic space. And the present experiment, with its demonstration of a different reference system for visual and kinesthetic constant errors, provides further evidence of the distinctiveness of the tactile-kinesthetic and visual representations. This is not, however, to deny intermodal effects, nor to preclude the postulation of some unitary spatial attributes under optimal stimulus conditions.

The experiment was conducted as part of a programme of research supported by the U.S. Air Force Office of Scientific Research, under grant, AFOSR 62–1.

## References

Kelvin, R. P. (1954). Discrimination of size by sight and touch. *Quart. J. exp. Psychol.*, **6**, 23–34.

Kelvin, R. P., and Mulik, A. (1958). Discrimination of length by sight and touch. *Quart. J. exp. Psychol.*, **10**, 187–92.

Künnapas, T. M. (1955). Influence of frame size on apparent length of a line. *J. exp. Psychol.*, **50**, 168–70.

KÜNNAPAS, T. M. (1957). The vertical-horizontal illusion and the visual field. *J. exp. Psychol.*, **53**, 405–7.

KÜNNAPAS, T. M. (1958). Influence of head inclination on the vertical-horizontal illusion. *J. Psychol.*, **46**, 179–85.

REID, R. L. (1954). An illusion of movement complementary to the horizontal-vertical illusion. *Quart. J. exp. Psychol.*, **6**, 107–11.

SANDSTRÖM, C. I. (1951). *Orientation in the Present Space.* Stockholm.

*Manuscript received 16th April, 1963.*

# APPARATUS

## NOISELESS ELECTRONIC SWITCH AND TIMER

BY

C. S. L. KEAY* and M. RODDA†

*University of Canterbury, Christchurch, New Zealand*

In psychoacoustic testing and experimentation the avoidance of extraneous stimuli is most important, but the elimination of such cues is often difficult. A circuit is described which enables sound stimuli to be switched on and off in a manner which produces no unwanted transient signals. The circuit may be adapted to function as a noiseless changeover switch.

### INTRODUCTION

A common source of difficulty in designing psychoacoustic experiments is the elimination of clicks introduced when circuits are "made" or "broken." These clicks form unwanted cues and unless they are removed they may result in erroneous conclusions being reached. The difficulty can be avoided by the use of an appropriate switching system. However, electrical and mechanical switching systems require very careful design to avoid the introduction of clicks or other irrelevant cues whenever the stimulus is changed. Even so, such systems may produce transient conditions and the resulting unwanted frequency components may act as extraneous stimuli.

The present circuit avoids these difficulties by changing the stimulus conditions gradually, over a period of approximately one sec. It also incorporates a timer circuit to simplify the experimenter's task in timing different stimuli. At the time of writing the unit has been in use for over a year in studying problems of pure-tone discrimination in rats and temporary shifts of threshold, resulting from prior auditory stimulation, in human subjects. Although the circuit described is only capable of on/off actions it may be converted to changeover operations by a simple modification.

### CIRCUIT DESCRIPTION

The circuit (see Fig. 1) basically consists of a single valve voltage amplifier stage, the amplification of which can be reduced to zero by means of a bias voltage applied to the suppressor grid: the same principle as that used by Barry and Degelman (1961). The valve chosen is a 6AU6 which has a very gradual (but complete) cut-off characteristic for its suppressor grid. The 0·5 microfarad capacitor attached directly to the suppressor grid does not permit rapid changes in the bias voltage at that point, with the result that abrupt cut-off (or switch on) of the signal through the valve is prevented, and clicks or any other unwanted switching transients are eliminated.

The operation of the amplifier is controlled by the key-switch S1A/B. In the down position the suppressor grid is grounded through a 1-megohm resistor, causing the amplifier to pass the input signal. In the central position of the key-switch the amplifier is controlled by the multivibrator timer circuit formed by the 12AU7 double-triode valve and its associated components. The timer switches the amplifier on and off for variable periods, the duration of which may be set by the potentiometer P2 and switch S2 (on duration) and the potentiometer P3 and switch S3 (off duration). For either condition the duration is continuously controllable from 1½ to 45 sec. with the appropriate switch closed and from 40 sec. to 6 min. with the switch open. Experience has shown that this range of time control is more than ample for all of the experiments described earlier.

The third (up) position of the key-switch S1A/B causes the amplifier suppressor to be returned to a negative voltage sufficient to cut the valve off (regardless of the state of the timer), thereby preventing any signal reaching the output. When the amplifier is deliberately switched on or off manually by the key-switch the timer continues to function, but without effect. At all times its state is displayed by the small neon indicator lamp in the multivibrator circuit: when the indicator is glowing the timer is in the 'on' state.

---

* Department of Physics.

† Department of Psychology.

Brief mention must be made of the coupling circuit between the multivibrator timer (12AU7) valve and the suppressor grid of the switched amplifier stage. When the left-hand triode section of the 12AU7 switches on, its anode voltage falls from earth potential to approximately 130 volts negative. This change is transmitted to the suppressor grid of the amplifier stage via the 1N1763 silicon diode. This cuts the amplifier off, but not instantaneously: the 0·5 microfarad capacitor $C_s$ tied to the suppressor grid takes time to charge to −130 volts through the diode and 33K resistor and as a result the signal passing through the amplifier takes just under one second to die away. This entirely eliminates clicks and sudden changes in signal level.

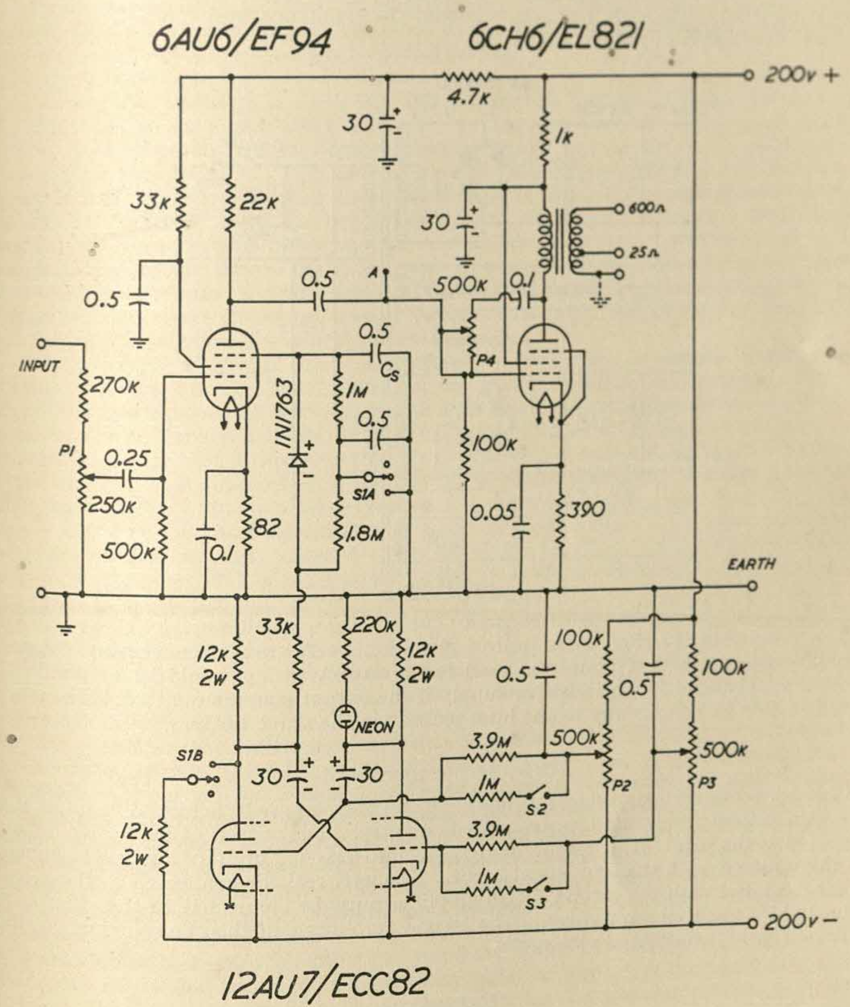


FIGURE 1

Circuit diagram of "Noiseless Electronic Switch and Timer."

When the left-hand triode section of the 12AU7 switches off the anode immediately returns to earth potential. This change would, if transmitted by the 1N1763 diode, cause the amplifier to switch on very quickly because the discharge rate of the capacitor $C_s$ is more rapid than the charge rate at the voltage level where the suppressor grid allows the amplifier stage to resume conduction. However, the 1N1763 diode is now non-conducting and the capacitor $C_s$ can only discharge through the resistor chain consisting of the 1M, 1·8M, 33K and 12K resistors. The 0·5 microfarad capacitor connected to earth between the first two resistors further reduces the initial discharge rate and the overall result is that the signal through the amplifier takes almost one second to build up to the desired strength. Although the distortion in the amplifier has been kept low throughout the audible frequency range a very small amount of harmonic distortion during the switching phase may be detected if an oscilloscope is used.

The output stage (6CH6 valve) was designed to feed a pair of calibrated earphones or provide a 600-ohm output to a calibrated attenuator. The potentiometer P4 is a treble-cut control. Other types of output circuit could just as readily be employed: the necessary signal being taken from the point label "A" in Fig. 1, and the circuitry to the right of that point is omitted. The only restriction is that the resistance from "A" to earth should remain in the region of 100,000 ohms.

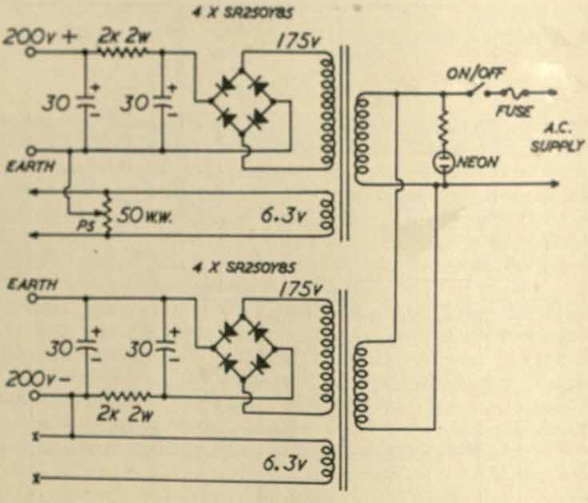


FIGURE 2

The power supply (Fig. 2) is quite conventional. A transformer rating of 40 mA secondary current at 175 volts was used but a 20 mA rating would be ample for the negative supply, and for the positive supply if the output stage is omitted. The pre-set potentiometer P5 cancels any mains hum picked via the valve heaters.

## Circuit Modification

Changeover operations, under automatic timer control, may be achieved by duplicating the 6AU6 amplifier valve and its associated circuitry. The output of the second 6AU6 is connected at the point "A," or, more preferably, into a second input of a separate amplifier and there mixed with the first signal input. The 33K resistor connected to the 1N1763 diode associated with the second 6AU6 amplifier must be connected to the other anode of the 12AU7 multivibrator valve. The second amplifier will then be switched on when the first is switched off, and vice versa.

## Reference

Barry, J. J. Jr. and Degelman, J. (1961). A high power noise amplifier with an electronic keying system. *J. exp. Anal. Behav.*, **4**, 85–6.

*Manuscript received 16th May, 1963.*

# BOOK REVIEWS

*The Behavioral Basis of Perception.* By James G. Taylor. New York and London: Yale University Press. 1962. Pp. xv + 379. 68s.

Mr. Taylor's general aim is to expound the role of learning in vision. Here he is relevant, forceful, and sometimes illuminating. His more specific aim is to present an elaborate deductive system, based on conditioning. This is laborious. It was developed, he says, "almost entirely without reference to the contributions of other workers in the field of perception." He need not have told us that.

The contents of the book are briefly as follows: Chapter one traces his indebtedness to Ashby and Hull, and says conditioning will be treated "not as a subject for inquiry or argument but as an explanatory principle." Chapters two and three use a notation derived from mathematical set theory (there are some 30 symbols to be memorized) to state the basic principles. This covers movements of the eyes, head and hands only. Chapters four to seven develop these principles to include movements of the trunk, adaptation to gravity, and the perception of size, shape and distance. Two chapters describe experiments with distorting spectacles: first, Stratton and Ewart, and a brief account of the Innsbruck studies (the figures on p. 175 seem to have been interchanged); then three good original experiments. After treating colour and the sensory modalities, Taylor criticizes other theorists, notably Gibson, Erismann, Wapner and Werner, Hebb, and Piaget. Finally: philosophical implications.

The gist of the theory is this. At birth we have, on the one hand, "many afferent impulses pouring into the brain," and, on the other, unlearned reactions (p. 23). The stimuli may be subdivided into drive stimuli, and stimuli affecting the receptors. The latter lead to S, defined as "the set of all possible states of the afferent system . . . as determined jointly by the environment, the orientation of eyes and head, and the position of the hand" (p. 40). S is described in terms of sets of elements. Take an infant, who must learn to reach for an object lying in a certain direction: "the apparently simple task of carrying the hand from an initial position Y to a goal at X requires at least 11 afferent variables" (p. 15), depending on the angle at which light enters the eye, the position of the eyes and head, and the initial position of the hand. The same object can produce many different afferent neural processes. This, writes Taylor, "gives rise to one of the central problems of perception. How do these widely differing elements . . . all give rise to the same perception?" (p. 26).

When the first correct movement of the hand occurs, the process of conditioning begins whereby the response is connected to the set of stimuli operating when the movement began. "Of course the successful response is not the only one that occurs, but the principles of conditioning enable us to deduce that those other responses will be eliminated, partly through extinction and partly because the probability that the successful response will be evoked immediately, thereby excluding all others, increases with every reinforcement" (p. 38). Conditioning gradually extends to all the sets of stimuli which precede the movement, so that, when these sets are presented again, the appropriate engram is aroused, and the organism is ready for the response (p. 42); the engram being, not a single chain of neurons, but a complex network of connections.

Ploughing through Taylor's labyrinth of symbols, I felt a sense of *déjà vu*. It was finally resolved at this reference (p. 127) to the constancies of shape and size: "It is clear now that there is a very simple explanation of these phenomena. No matter what distortions the retinal image of an object may suffer as a result of its position relative to the eyes, the final response of handling the object has certain components that remain invariant from one occasion to another." Compare Berkeley (1709): " . . . the visible object still changing as your approach to or recede from the tangible object, it hath no fixed and determinate greatness. Whenever therefore we speak of the magnitude of any thing . . . we must mean the tangible magnitude; otherwise there can be nothing steady and free from ambiguity spoken of it." The similarity between the principles of Conditioning and Association has often been noted, and Taylor's views seem to resemble closely the venerable empirical tradition of Berkeley, and Mills and Bain. Behind the imposing panoply of Hullian technicalities stands the quiet figure of the Bishop of Cloyne.

This historical parallel points to the book's main strength and weakness. With Gestalt psychology, the empirical tradition was disrupted. Now it is reviving, notably in Hebb. Taylor's re-statement of it is also valuable. If he had placed it in its historical context, it would have gained in perspective. Even so, the experiments with distorting spectacles seem to provide one of the most interesting sources of evidence to date.

But associationism was attacked, first by Ward and Stout, then by Gestalt psychologists. A revival of empiricism must remember this. How does Taylor's version deal with the difficulties they raised? Gestalt psychology is almost wholly neglected. In a book asserting empiricism, and purporting to explain the constancies, this may fairly be termed astonishing. (There is also no reference to Ames, Boring, Brunswik, Katz, Kilpatrick, Ittelson, Lambercier, Thouless and Wallach. Indeed, he says he "took the deliberate decision to read no more of the current literature on perception.") Many points might be raised here, but the basic objection to associationism was its atomism or elementarism, and I am not sure that Taylor escapes this. He defines the sets of variables as "collections of elements" (p. 7). As I understand him, the element is ultimately the cortical terminus of an afferent neuron; it is this which, together with the other elements of the set, is linked to the response (p. 28). Now, on different occasions a given element would appear to belong to different sets, and is therefore linked with different responses. Which connection is excited on a given occasion must depend on which set the element belongs to. But is the element supposed to know which set it belongs to? Where does this crucial item of information come from? As the Gestalt psychologists would have said, the pattern or context must be considered. But perhaps I have missed Taylor's explanation.

This is not to say that Gestalt theory was correct in neglecting learning. As Stout remarked, they threw out the baby (empirical explanation) with the bathwater (associationism, elementarism). It is unfortunate that Taylor seems bent on recovering the bathwater; but his attempt to rescue the baby is an important one, and deserves close attention.

R. B. JOYNSON.

*The Brain as a Computer*. By F. H. George. Oxford: Pergamon Press. 1961. Pp. 413. 63s.

In some ways the problem of understanding effectively how the brain works resembles the problem that faced the early engineers concerning the effective action of steam. If we trace the action of one neurone or one water molecule we get nowhere, but a calculus of organization could give insight. For the steam engineer the effective study was thermodynamics, for the brain engineer it will be the sister science of cybernetics.

Dr. George in the first half of his book describes and explains the equivalence of several distinct aspects of cybernetics, namely (i) information theory, (ii) automata finite and infinite, (iii) logical nets, (iv) specially programmed digital computers and (v) hardware models.

This part of the book is not easy to grasp, for it seeks to compress into a limited compass some rather sophisticated mathematical logics developed in a distinctly formidable symbolism, not all of which is explained. The reader nevertheless will probably emerge convinced that the language of logical sequences may be expressed in symbols or models of the utmost variety, that these languages are roughly equivalent, and that each is rich enough to go a long way towards describing the logical structure of the computer-like reactions of the brain.

Now the effectiveness of a method whether analytical or experimental is generally seen in its application, and the second half of the book applies cybernetics to some questions of experimental psychology and physiology. In this Dr. George has been no more successful than his predecessors, for no light is thrown upon any brain problem by any form of cybernetic.

Modern concepts of biological mechanism of course are dominated by the general cybernetic idea of goal-seeking, but the mathematics of strict cybernetics are so hard to apply effectively to actual psychology that most analysts make no attempt to do so. They prefer to invent bogus neurone networks with regular, and artificial properties and to use the names "learning," "recognition," "memory," etc. for the contrived performance.

Unlike some of these, Dr. George knows very well the important experimental work in psychology and neurophysiology, and he discusses difficulties and alternative views, especially in Learning Theory to which he devotes 50 pages. What, however, is discouraging in his treatment is not so much the failure to use cybernetics to settle anything that had not been settled already, as the lack of any indication of how cybernetics *could* be usefully applied. The models themselves are so flexible that they may be made to imitate anything you find. But models that can "explain" everything can predict nothing. Hardware is entertaining rather than illuminating. If cybernetics is going to give us insight into the organization of the brain surely it must do better than record or reproduce merely what we know already.

We certainly need the two halves of Dr. George's book, the language of organization and the harvest of systematic experimental observation, but we need something more—a method of so combining them that new principles emerge—the principles that underly the actual organization of the brain. Then at last models will have some predictive value for, by the rules of their construction, they will logically represent the Brain as a Computer.

W. A. H. Rushton.

*Recent Advances in Biological Psychiatry, Vol. IV.* Edited by Joseph Wortis, M.D. New York: Plenum Press. 1962. Pp. xiii + 385. $12.50. £5 8s.

The title of this volume is perhaps misleading but the area it covers is explicitly described in the editor's introduction. Firstly it presents a picture of advances in experimental psychiatry "as they are reflected in the papers presented" at the "sixteenth annual convention of the Society of Biological Psychiatry" in 1961, and also records "symposia of timely interest" which took place at the meeting. The contents include a record of a joint meeting of the Pavlovian Society and the Society of Biological Psychiatry—nine papers—three papers on ethology, 16 papers in the field of psychopharmacology, and four miscellaneous papers which include a useful review of the phenomena elicited by double simultaneous stimulation and a review of some of the biochemical theories of schizophrenia. As might be expected there is something for everyone but the standard is uneven.

Probably American psychologists and experimental psychiatrists lead the world in their ability to publish the same material in different journals. If essentially the same work is also to be published in the proceedings of every society to which each member of the research team belongs, then the problem facing those who wish to keep track of current work in their own field will become even more difficult than it is at present. However, several papers in the present volume are well worth reading and it is clearly preferable that the proceedings of a learned society should be published in full rather than as brief didactic notes. If, as with some similar publications in this country, economic pressures mean authors publishing in this way are denied offprints, then the trend is clearly to be deplored.

Alick Elithorn.

*Speech and Thought in Severe Subnormality.* By N. O'Connor and B. Hermelin. Oxford: Pergamon Press. 1963. Pp. xi + 122. 25s.

*Signs, Signals and Symbols.* A presentation of the British approach to Speech Pathology and Therapy. Ed. Stella E. Mason. London: Methuen. 1963. Pp. xii + 212. 35s.

With *Speech and Thought in Severe Subnormality* one welcomes with pleasure a book by clinical psychologists describing experimental investigations instead of the usual standardised test results. By comparing the behaviour of severely subnormal individuals with that of normal children of the same mental age on small discrete tasks, it is pointed out that there are many functions in which the two groups differ radically. Each separate investigation may not in itself appear to add up to much but the total weight of evidence produced gives one hope that a greater understanding may soon be possible of those functions commonly called intelligence. Marked differences in the performance of different subnormal sub-groups are also demonstrated illustrating, as the authors point out, the danger of regarding intelligence as a single function and measuring it in global terms such as I.Q.

As a result of this work, many of our most cherished beliefs may have to be revised. Imbeciles apparently do not perceive worse in all modalities than normal people (their stereognostic perception seems to be somewhat better), nor do they forget what they have learned adequately more quickly. Moreover, the use of words does not apparently always assist adaptation. It is sobering to discover that imbeciles learned a reversal-learning task quicker than normals until they were encouraged to use words during the original learning period. After this, their ability to reverse was slowed up to the same speed as normals.

As a result of this and similar findings, the authors suggest that one of the differences between the thinking behaviour of normal and subnormal individuals is that normal people single out the relevant features in a display, label them with names and then operate within the verbal code. Such verbal coding does not occur spontaneously in imbeciles.

This book indicates that many of our favourite catch-words (e.g. "concrete" and "abstract") may have to be abandoned. It would be a pity if they were replaced too soon by other catch-words more currently OK (e.g. "coding") but so long as attention continues to be focused as it is here on the conditions and variables affecting behaviour rather than on hypothetical models designed to explain it, there is hope for the future.

In *Signs, Signals and Symbols* we are introduced to another side of the picture; namely the vast range of human activities with which speech is concerned and the number of different disciplines which have to be entered for an adequate appreciation of its importance. The book consists of papers read by members and their guests at a National Conference of the College of Speech Therapists held in 1961. It skirts superficially over the history of linguistics, describes some of the latest inventions for measuring phonation and hearing and includes a series of accounts by Speech Therapists themselves of cases (predominantly stammerers) under treatment (predominantly psychoanalytic). None of the speakers could go deeply into their subjects in the space and time allotted but as a general survey of the field and of the problems involved in dealing with speech-disordered people, it is of value.

MOYRA WILLIAMS.

*Thinking: An Introduction to its Experimental Psychology.* By George Humphrey. London and New York: Wiley, Science Editions. 1963. Pp. xi + 331. 17s.

First published in 1951, this well-known book was reviewed by Sir Frederic Bartlett (who first suggested its writing to the author) in this *Journal*, 1952, **4**, 87–90. While the new and cheaper edition is very welcome, many will regret that Professor Humphrey has not revised his book to take account of relevant work carried out during the past 12 years, e.g. Bruner's work on conceptual "strategies" and the work of Luria and many others on language and communication. None the less, Professor Humphrey's account of the traditional experimental psychology of thinking, in particular the Denkpsychologie of the Würzburg School and Selz, remains by far the best available.

O. L. ZANGWILL.

*Decisions, Values and Groups.* Edited by Norman F. Washburne. Oxford: Pergamon Press. 1962. Pp. 521. £5.

An Interdisciplinary Behavioral Sciences Research Conference was held in 1958 and out of the papers and discussions that then took place, emerged the articles of which this volume consists. A wide range is covered. Learning-, communication-, decision- and organization-theory are all represented in papers by experimental, clinical and social psychologists, by political scientists, sociologists and social anthropologists. The bias is markedly towards conceptual analysis, on occasion right up to logical formalization of theory. The editing of the book is minimal and to read it is like reading a volume of Behavioral Science straight through, except that the standard of the papers is higher and some very long ones are included. One aim of the book was to encourage by its example cross-fertilization of disciplines, but few of the writers have made any concessions for the sake of communication to non-specialists in their field. There seems to be some connection between inability to make clear, concern with definitions and co-operation in Babel-like projects. But individually considered, at the level at which one can understand, there are some good papers here.

W. H. N. HOTOPF.

*An Introduction to the Science of Human Behavior.* By J. I. Nurnberger, C. B. Ferster and J. P. Brady. Foreword by D. McK. Rioch. New York: Appleton-Century-Crofts. 1963. Pp. xvi + 379. $5.95.

This highly competent book is divided into three distinct sections. Dr. Brady writes on the biological substrata of behaviour, Dr. Nurnberger on its psychosocial determinants, and Dr. Ferster on behaviour from the standpoint of a Skinnerian. In his Foreword, Dr. Rioch expresses the hope that this volume "represents a major step in bridging the conceptual gap between the laboratory and the clinic." Maybe.

O. L. ZANGWILL.

*Electrical and Drug Treatments in Psychiatry.* By A. Spencer Paterson. Amsterdam and London: Elsevier. 1963. Pp. xii + 248. 56s.

This is a clear, straightforward and up to date account intended for psychiatrists in practice. It will no doubt also prove informative to clinical psychologists and others who have to do with psychiatric patients under treatment. To the merely curious, the scale and sophistication of modern psychopharmacology may come as a distinct surprise.

O. L. ZANGWILL.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XVI 1964 Part 4

# THE EFFECT OF SOMAESTHETIC AND ACOUSTIC STIMULI ON THE THRESHOLD OF FUSION OF PAIRED LIGHT FLASHES IN HUMAN SUBJECTS

BY

G. HORN and P. H. VENABLES

*From the Department of Anatomy, University of Cambridge, and The Social Psychiatry Research Unit, Institute of Psychiatry, London*

The threshold of fusion of paired flashes of light was measured by determining the largest interval between two flashes of light at which they were reported as one and not as two flashes. When a weak electric shock to the skin was presented at the same time as the first flash, the threshold was increased compared to the threshold measured when the flashes were unaccompanied by a shock. As the interval between shock and first flash was increased up to 600 millisec., the effect diminished. A similar pattern of results was obtained using a click, instead of a shock as the additional stimulus. The effect on the two-flash threshold of varying click intensity was also studied. No retroactive effects of the shock on the two-flash threshold were observed.

## INTRODUCTION

The present series of experiments were undertaken as the result of certain observations made in the course of a study of single cells in the visual cortex of the unrestrained cat (Horn, 1963). Briefly, these observations were that the impulse activity of some cells was modified by a single weak electric shock applied to the skin over the animal's back, as well as by diffuse flashes of light. The latency of the cellular response following a shock was, on average, similar to that following a flash. Furthermore, if the shock was made to coincide with, or to precede the flash by an interval of up to 590 millisec., the response of the cell to the flash was changed compared with the response when the flash was unaccompanied by the shock. Thus, signals evoked by somaesthetic and photic stimuli appeared to interact in the visual pathways.

It was considered that these interactions, observed in animals, might have their counterpart at the psychophysical level in human subjects; that a weak shock to the skin, if appropriately timed, might interfere with the performance of a visual discrimination task. It remained uncertain, however, whether or not the results obtained were due to the emotionally disturbing effects of shock; so a further series of experiments was conducted using an acoustic "click" as the additional stimulus.

## METHOD

### (*a*) *Visual task*

The subjects sat in front of and approximately three feet away from a cathode ray tube, the face of which could be diffusely illuminated. The intensity of each flash was 200 foot-lamberts with a duration of 5 millisec. Two flashes were delivered, one flash following the other, and the interval between them could be varied by the experimenter.

The subject was required to press a button once if he saw one flash and twice if he saw two flashes.

The threshold of fusion was determined by the following technique. A pair of flashes with a long inter-flash interval (I.F.I.) was presented; if the subject pressed twice, flashes with a short I.F.I. were presented; if the subject then pressed once the procedure was repeated with decreasing range until two I.F.I.s a millisec. apart were reached. Flashes at these I.F.I.s were then twice repeated and if the responses were maintained the I.F.I. at which the subject pressed once was taken as the threshold.

The following sequence is from a typical record:

100–2, 50–1, 90–2, 60–1, 70–1, 80–2, 70–1, 75–1, 78–2, 75–1, 76–1, 76–2, 75–1, 75–1, 76–2, 75–1, 76–2; i.e. thresholds of all subjects could be measured to the nearest millisec. As the reliability of such a short procedure might be in doubt, in a preliminary experiment thresholds were determined on 58 subjects in two periods four minutes apart. A coefficient of 0·96 was obtained (Venables, 1963).

This two-flash threshold discrimination task was used as the visual task in all the experiments described in this study, in each of which different groups of subjects were used.

### (*b*) *Electric shock*

The shock was delivered through a pair of electrodes placed approximately 2½ in. apart on the ventral surface of the forearm (the button for signalling the number of flashes was held in the opposite hand). The electrodes which were of silver were used with electrode jelly (Cambridge Instrument Company), which had been previously rubbed into the sites. Collodion cement was applied around the periphery of the electrode to prevent the electrolyte drying out.

The shock consisted of a square pulse 10 millisec. in duration, delivered from a Tektronix Pulse Generator having a maximum output impedance of 1800 ohms. The amplitude of the pulse was 10 V. greater than threshold and ranged from 12 to 30 V. Similar pulse characteristics were used to deliver shocks to the cats in earlier experiments (Horn, 1963).

### (*c*) *Acoustic clicks*

Clicks were delivered through Brown type F earphones energized by square wave pulses delivered from the Pulse Generator used for presenting shocks. The parameters of the pulse used were 1 millisec. duration and either 5 V. or 40 V. amplitude. Subjects' thresholds for clicks were measured approximately by using 0·1 millisec. pulses and determining the amplitude of voltage which produced a just audible click. This measurement was used as a screening procedure to eliminate any subjects with defective hearing rather than to obtain an absolute threshold.

The time interval between the first of the paired light flashes and the preceding shock or click stimulus was controlled using Tektronix Pulse and Waveform Generators.

The experiments were conducted in a room darkened except for a screened light which illuminated the timing apparatus used by the experimenter.

## RESULTS

### *Experiment 1*

The aim of this experiment was to determine the effect on two-flash threshold of shocks presented 0, 200, 400 and 600 millisec. before the first of each pair of flashes. The order of presentation of shock-flash interval was varied for each subject according to the letters in the row of a Latin Square. The two-flash threshold, measured in millisec., was first determined without shocks. The subject was then warned that he might receive shocks, and his two-flash threshold was again measured, though no shocks were in fact given. The next four thresholds were determined, the flashes being associated with shocks in the order described above. Immediately after this series was complete a second threshold without shock was measured. The whole procedure took about 20 min. for each subject. Twelve subjects took part, all were medical students. They were told that the experiment was "concerned with vision" and that they would receive weak electric shocks to the skin. They were asked to disregard the shocks and to report whether they saw one or two flashes of light.

Analyses of variances were carried out separately on the data (*a*) for those thresholds where the flashes were associated with shocks, and (*b*) for those thresholds where the flashes were not accompanied by shocks. No difference was found between the first threshold (mean 62·1, S.D. 9·8 millisec.) and that determined under "threat of shock" (mean 64·6, S.D. 10·7 millisec.). Both of these, however, differed ($p < 0·01$) from the final no-shock threshold (mean 73·7, S.D. 14·1 millisec.) at the end of the series.

TABLE I

Latin Square Analysis of Variance of Thresholds at Four Shock flash Intervals

| *Source* | *d.f.* | *M.S.* | *F* | *P* |
|---|---|---|---|---|
| Shock-flash interval .. .. | 3 | 609·28 | 13·61 | < 0·001 |
| Orders .. .. .. .. | 3 | 122·39 | 20·73 | N.S. |
| Subjects .. .. .. .. | 11 | 790·88 | 17·60 | < 0·001 |
| Residual .. .. .. .. | 30 | 44·76 | — | — |
| Total .. .. .. .. | 47 | — | — | — |
| Distribution of thresholds: | | | | |
| Shock-flash interval (millisec.) | 0 | 200 | 400 | 600 |
| Mean ± standard deviation threshold (millisec.) .. | 83·0 ± 17·9 | 76·1 ± 14·1 | 69·2 ± 14·4 | 67·3 ± 11·9 |

Analysis of the thresholds measured when the flashes were preceded by shock (Table I) showed a highly significant term for the factor of shock-flash interval. The threshold, measured when shock and first flash were simultaneous, differed significantly from thresholds measured at all other shock-flash intervals; that measured at a shock-flash interval of 200 millisec. differed from those at 400 and 600 millisec. but these latter two thresholds did not differ. These results are shown graphically in Figure 1.

The threshold for the zero shock-flash interval is greater ($p < 0·01$) than the final no-shock threshold, and that for the 600 millisec. interval is less ($p < 0·05$) than this final threshold.

### *Experiment* 2

Retroactive effects of an additional stimulus on the absolute threshold to a test stimulus have been described by Howarth and Treisman (1958) and Halliday and Mingay (1961). To test for this effect the procedure was altered so that the shock *followed* the first flash by 50, 100, 150 or 200 millisec. No-shock thresholds were measured at the start and at the end of the session. The four flash-shock intervals were arranged in Latin Square order. Four subjects were tested.

At the 50 millisec. flash-shock interval the threshold was significantly ($p < 0·05$) higher than at any of the other flash-shock intervals. Thresholds at flash-shock intervals of 100, 150 and 200 millisec. did not differ among themselves (overall mean 73·9, S.D. 11·9 millisec.) and did not differ from those thresholds unaccompanied by shock (overall mean 73·0, S.D. 9·5 millisec.) which did not differ between themselves. At the 50 millisec. flash-shock interval, the shock actually preceded the second flash (mean threshold 95·3, S.D. 24·2 millisec.) and is thus comparable to the zero shock-flash interval of Experiment 1.

These two experiments suggest that when a shock precedes the first of a pair of flashes of light by less than 200 millisec. the capacity to discriminate between the two flashes is impaired. There is also a suggestion that performance, relative to the final no-shock threshold, is improved when a shock precedes the first flash by 600 millisec. No retro-active effects were observed.

*Experiment* 3

The aim of this experiment was to determine the effect of clicks on two-flash threshold, when these clicks were presented as in Experiment 1, at 0, 200, 400 and 600 millisec. before the first of each pair of flashes. The procedure was similar to that employed in Experiment 1 except that no "threat of click," analogous to the "threat of shock," threshold was measured.

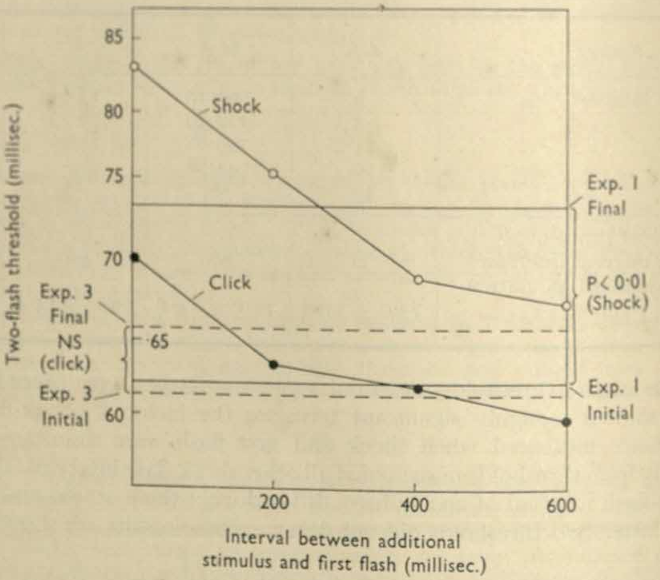


FIGURE 1

Two-flash threshold for varying shock-flash intervals (Experiment 1, open circles) and click-flash intervals (Experiment 3, closed circles).

The eight subjects tested were medical students and members of the Department of Anatomy. Throughout this experiment the stimuli were presented through earphones to which were delivered pulses of 5 V. amplitude and 1·0 millisec. duration.

Analyses of variance were again carried out separately for the data gathered in Latin Square form and also for those thresholds measured when flashes were not preceded by clicks.

No difference was found between the first no-click threshold (mean 61·3, S.D. 9·5 millisec.) and the final no-click threshold (mean 65·6, S.D. 8·1 millisec.). The results of analysis of variance are shown in Table II. It may be seen that, as with Experiment 1, there is a highly significant term for the factor of click-flash interval.

The threshold given at zero click-flash interval differed significantly from the threshold at all other intervals, that at 200 millisec. differed from that at 600 millisec., but no other differences were significant. These results are plotted in Figure 1.

The threshold for zero-click-flash interval is greater ($p < 0{\cdot}05$) than the final no-click threshold, which in turn is greater ($p < 0{\cdot}05$) than that for the 600 millisec. interval.

TABLE II

LATIN SQUARE ANALYSIS OF VARIANCE OF THRESHOLDS AT FOUR CLICK-FLASH INTERVALS

| *Source* | *d.f.* | *M.S.* | *F* | *P* |
|---|---|---|---|---|
| Click-flash interval | 3 | 160·54 | 10·77 | < 0·001 |
| Orders | 3 | 9·04 | 0·61 | N.S. |
| Subjects | 7 | 545·41 | 36·60 | < 0·001 |
| Residual | 18 | 14·9 | — | — |
| Total | 31 | — | — | — |
| Distribution of thresholds: Click-flash interval (millisec.) | 0 | 200 | 400 | 600 |
| Mean ± standard deviation threshold (millisec.) | 70·1 ± 12·0 | 63·5 ± 11·1 | 61·9 ± 11·7 | 59·8 ± 10·7 |

It will be seen from Figure 1 that although the mean initial thresholds of the subjects in Experiments 1 and 3 are similar, the thresholds measured when the flashes were associated with shocks are, on the whole, greater than those associated with clicks. These two sets of data were examined in more detail in the following way. From the 12 subjects in Experiment 1, eight were selected whose initial thresholds most closely matched those of the eight subjects in Experiment 3. Latin Square order was preserved and an analysis (Table III) comparing the two experiments was then undertaken.

TABLE III

LATIN SQUARE ANALYSIS OF VARIANCE OF THRESHOLDS OF MATCHED SUBJECTS FROM EXPERIMENTS 1 AND 3

| *Source* | | *d.f.* | *M.S.* | *F* | *P* |
|---|---|---|---|---|---|
| Shock v. Click | | 1 | 2562·89 | 91·2 | < 0·001 |
| Shock/click interval | | 3 | 624·93 | 22·2 | < 0·001 |
| Interaction | | 3 | 60·89 | 2·16 | N.S. |
| Shock | Subjects | 7 | 940·88 | 33·47 | < 0·001 |
| | Order | 3 | 126·61 | 4·50 | < 0·01 |
| Click | Subjects | 7 | 545·41 | 19·40 | < 0·001 |
| | Order | 3 | 9·04 | 0·32 | N.S. |
| Residual | | 36 | 28·11 | — | — |
| Total | | 63 | — | — | — |
| Distribution of thresholds: Interval (millisec.) | | 0 | 200 | 400 | 600 |
| Mean threshold Shock-flash (millisec.) | | 86·6 | 78·8 | 72·6 | 67·9 |
| Mean threshold Click-flash (millisec.) | | 70·1 | 63·5 | 61·9 | 59·8 |

The mean initial no-shock threshold was 62·8, S.D. 12·3 millisec., and the mean initial no-click threshold was 61·3, S.D. 9·5 millisec. The mean final no-shock threshold (73·6, S.D. 13·7 millisec.) differed significantly ($p < 0·05$) from the mean final no-click threshold (75·6, S.D. 16·1 millisec.). The results of the analysis shown in Table III confirm that the shock-flash thresholds are all significantly greater than the click-flash thresholds. The fact that the interaction term is not significant shows that the size of the threshold differences is effectively constant at all intervals.

This difference is of the same order of magnitude as the difference between the mean final no-shock threshold and the mean final no-click threshold. Whether this difference results from stimulating different modalities, or is due to differences in relative intensity of the two interfering stimuli, cannot be determined from the present data.

The results of Experiment 3 suggested that intensity may be an important parameter affecting the results. The interval range between 0 and 200 millisec. unexamined in either Experiments 1 or 3 also required investigation. This experiment was therefore carried out with two intensities of click; 5 V. as in Experiment 3 and 40 V., with click-flash intervals of 0, 75 and 150 millisec. Twelve subjects were used. all of whom were medical students. The 40 V. clicks were uncomfortably loud and several trial clicks were needed to allow the subject to become adapted to their intensity. If he blinked as a result of the click he was asked to report and that threshold determination was excluded. The subject was warned of the change of intensity.

The interval and intensity combination were presented as before in the row order of a Latin Square to each subject. No-click thresholds were taken at the beginning and end as before. The mean initial threshold was 63·8, S.D. 9·9 millisec. and the mean final threshold was 67·1, S.D. 11·2 millisec. These two values did not differ significantly. Analysis of variance was carried out on the click-flash threshold data and is presented in Table IV and the data plotted in Figure 2.

TABLE IV

Latin Square Analysis of Variance of Click-flash Thresholds at Three Intervals and Two Intensities

| *Source* | *d.f.* | *M.S.* | *F* | *P* |
|---|---|---|---|---|
| Treatments | 5 | 1708·92 | 11·04 | < 0·001 |
| Order | 5 | 15·01 | 0·09 | N.S. |
| Subjects | 11 | 1908·27 | 12·33 | < 0·001 |
| Residual | 50 | 154·67 | — | — |
| Total | 71 | — | — | — |
| Distribution of thresholds: | | | | |
| Click-flash interval (millisec.) | | 0 | 75 | 150 |
| Mean ± standard deviation threshold 5 V. intensity (millisec.) | | 84·6 ± 23·1 | 75·2 ± 14·0 | 68·9 ± 13·4 |
| Mean ± standard deviation threshold 40 V. intensity (millisec.) | | 99·3 ± 29·2 | 87·7 ± 20·6 | 67·0 ± 14·6 |

Subsequent *t* tests show that there are significant differences in thresholds between intensities at click-flash intervals of 0 and 75 millisec. Thresholds at all the intervals at 40 V. intensity differ significantly from each other. At 5 V. intensity only thresholds at 0 and 150 millisec. intervals differ from each other. Both zero interval thresholds and the 40 V., 75 millisec. interval threshold differ from the mean of the no-click thresholds, but no other thresholds differ from it.

On one hand, therefore, the presence of a shock increases the two-flash threshold above that resulting from the presence of a click, for all intervals. On the other hand, increasing the intensity of the click appears to affect only the short click-flash intervals. The pattern of the experiments using two-click intensities is thus different from that resulting from the experiments using two modalities.

## Discussion

The experiments which have been described have yielded results which appear to bear an interesting relationship to the neurophysiological studies on the cat (Horn, 1963). In these it was found that the effect of shock on the response of a cell to a flash of light varied with the interval between the two stimuli. That is when the shock coincided with the flash, cells tended, on average, to fire less vigorously than when the flash was presented alone; when the shock preceded the flash by between 300 and 600 millisec. there was some indication that the response to the flash was, on average, slightly greater than when the flash was unaccompanied by the shock. These results were obtained on a small sample of cells and it would be unwise to put too much store by the actual magnitude of the changes observed. In any case, there was very considerable variation from cell to cell. The important thing, from the point of view of the present study, is that shock interfered with the response of some cells in the striate cortex to a photic stimulus, and this interfering effect changed as the shock-flash interval was varied. In the psychophysical studies which have been described, shocks interfered with the visual task and the magnitude of the effect varies with the shock-flash interval.

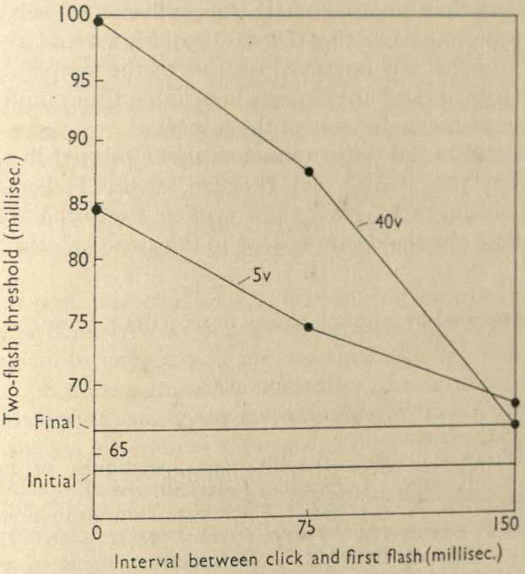


FIGURE 2

Two-flash thresholds for click-flash intervals at two click intensities.

Lindsley (1958) and Steriade and Demetriscu (1962) have shown that electrical stimulation of the reticular formation determines, in some way, the appearance of one or of two evoked potentials at the cerebral cortex in response to two stimuli presented in rapid succession. These experiments provide rather indirect evidence that the reticular formation may play some role in mediating the effects that have been described in our work.

Since similar effects on the discrimination of two flashes of light were obtained by clicks as well as by shocks it is important to know whether the cells in the cat's visual cortex also respond to click stimuli. There is evidence that acoustic stimuli modify activity of cells in the striate cortex of rabbits (Lömo and Mollica, 1962) and Dumont and Dell (1958) have found, in cats whose spinal cord had been transected at the cervical level ("encéphale isolé"), that the slow wave response of the visual cortex to electric shocks applied to the optic chiasma was modified by an acoustic stimulus.

When the additional sensory stimuli occur less than 200 millisec. before the first of two flashes, discrimination of them is impaired; but when the interval is 600 millisec., there is a suggestion that it is improved relative to the final two-flash threshold. The temporal sequence of these results bears a close parallel to those of Davis (1959) on the "psychological refractory period." If a subject is required to respond to the second of two stimuli (visual or auditory), the reaction time is prolonged, relative to the simple reaction time to a single stimulus, when the interval between the two stimuli is less than approximately 200 millisec. When the interval was 500 millisec., there was some suggestion (Davis, 1959, Figs. 1 and 2), that the reaction time to the second stimulus was improved relative to the simple reaction time. In these experiments it appears that the changes in reaction time could not be accounted for as a delay arising on the motor side of the response.

The results presented in this paper cannot easily be related to the previous work on heteromodal interaction (Gregg and Brogden, 1952; Treisman and Howarth, 1959; Halliday and Mingay, 1961) since these workers measured absolute thresholds. It is not clear how such thresholds are related to the two-flash threshold used in the present studies.

This work was supported by a grant to one of us (G.H.) from the Medical Research Council.

## REFERENCES

DAVIS, R. (1959). The role of "attention" in the psychological refractory period. *Quart. J. exp. Psychol.*, **11**, 211–20.

DUMONT, S., and DELL, P. (1958). Facilitations specifiques et non-specifiques des réponses visuelles corticales. *J. Physiol., Paris*, **50**, 261–4.

GREGG, L. W. and BROGDEN, W. J. (1952). The effect of simultaneous visual stimulation on absolute auditory sensitivity. *J. exp. Psychol.*, **43**, 179–86.

HALLIDAY, A. M., and MINGAY, R. (1961). Retroactive raising of a sensory threshold by a contralateral stimulus. *Quart. J. exp. Psychol.*, **13**, 1–11.

HORN, G. (1963). The response of single units in the striate cortex of unanaesthetized cats to photic and somaesthetic stimuli. *J. Physiol.*, **165**, 80 P.

HOWARTH, C. I., and TREISMAN, M. (1958). Lowering of an auditory threshold produced by a light signal occurring after the threshold stimulus. *Nature*, **182**, 1093–4.

LINDSLEY, D. B. (1958). The reticular system and perceptual discrimination. *In* JASPER, H. H., PROCTOR, L. D., KNIGHTON, R. S., NOSHAY, W. C., and COSTELLO, R. T., Eds. *The Reticular Formation of the Brain.* London.

LÖMO, T., and MOLLICA, A. (1962). Activity of single units in the primary optic cortex of the unanaesthetized rabbit during visual, acoustic, olfactory and painful stimulation. *Arch. ital. Biol.*, **100**, 86–120.

STERIADE, M., and DEMETRISCU, M. (1962). The reticular facilitation of responses to acoustic stimuli. *Electroenceph. clin. Neurophysiol.*, **14**, 21–36.

TREISMAN, M., and HOWARTH, C. I. (1959). Changes in threshold level produced by a signal preceding or following the threshold stimulus. *Quart. J. exp. Psychol.*, **11**, 129–42.

VENABLES, P. H. (1963). The relationship between level of skin potential and fusion of paired light flashes in schizophrenic and normal subjects. *J. Psychiat. Res.* **1**, 279–87.

*Manuscript received 6th May, 1963.*

# A CHRONOMETRIC STUDY OF SOME RELATIONS BETWEEN SENTENCES

BY

GEORGE A. MILLER and KATHRYN OJEMANN McKEAN

*From the Center for Cognitive Studies, Harvard University*

An attempt was made to study how people handle syntactic relations among English sentences by measuring the time required to convert one type of sentence into another. Preliminary results based on rate of work in a pencil-and-paper test are briefly summarized. More detailed data are given for an experiment in which subjects controlled the duration of presentation of the sentence they transformed. Presentation times were measured and compared with the presentation times when no transformation was required; the difference between these times was taken as a measure of the additional time required to perform the transformation. In such tasks, an active-passive difference requires more additional time than does an affirmative-negative difference; when both are involved the additional time required is approximately the sum for the two separately. Differences in verb constructions, however, all require about the same amount of time, and no additive relation is apparent. The relation of these results to the distinction between rules of formation and rules of transformation in descriptive linguistics is discussed.

A central task of psycholinguistics is to account for the fact that the people who know a language can deal with an unlimited variety of novel utterances in it. On the basis of a very finite exposure, a normal child quickly acquires an ability to understand and to produce admissible utterances that are different in form and content from any he has ever heard or spoken before. This ability to project from a finite corpus of actual utterances into an infinite set of potential utterances is what we should expect our psycholinguistic theories to explain.

If one begins this explanation, as many psychologists have, with the assumption that speech is a kind of response, then it is natural to assimilate to human vocal responses all those general attributes that responses are known to have in other psychophysiological contexts. In one version, for example, the principles of conditioning theory are brought to bear; words are treated as conditioned reflexes under the control of more or less the same variables that affect conditioned reflexes everywhere. According to conditioning theory, of course, particular experiences are most simply projected into novel situations by stimulus generalization. A more complicated kind of projection is presumably possible if the order of elements in a chained reflex can be varied; unfortunately, this kind of combinatorial projection, which is so extremely important in human communication, is less well explained by conditioning theory. In any case, it is by no means obvious that stimulus generalization and response chaining are adequate to carry all the theoretical burdens imposed by human language; even the most optimistic advocates of this approach must recognize a considerable gap between their explanatory principles and the phenomena of language.

One attempt to close this gap has resulted in the consideration of abstract types of similarity, analogous presumably to the kind of perceptual similarity that underlies stimulus generalization; thus we have heard about semantic generalization, propositional generalization, contextual generalization, etc., which are all suggestive terms, but need considerable refinement and clarification before they can explain the phenomena of language.

An alternative approach to the psycholinguistic problem begins by considering the nature of rules and rule systems. Most people believe that human language is governed by rules; phonological, semantic, functional, syntactic, and perhaps other types of rules are often distinguished. If one assumes that a language user knows, not responses, but rules for interpreting and producing linguistic utterances, then the unlimited variety of admissible utterances seems less surprising. (It merely indicates that some rules can be applied indefinitely often, i.e. "recursively," in the formation of admissible utterances.) This approach, emphasizing rules rather than responses, also raises questions concerning similarity. In the most general terms, linguistic rules are social conventions of the form, "When the same situation occurs again, do the same things." Such statements would be meaningless, of course, if we could not recognize a new situation as similar to an old one, or if we did not know what it meant to do the same things. Each rule must be supported by a system of recognizable similarities and differences. In these terms, even stimulus generalization could be viewed as the result of applying a rule—a highly adaptive rule that presumably does not need to be learned in most species. The more abstract types of similarity that underlie most of our linguistic rules, however, seem much more complicated and specialized.

In the present research a sentence would be presented and a subject asked to respond with another sentence that bore some specified syntactic relation to it. If a sentence of type S was presented, he was to transform it into one of type S′; if one of type S′ was presented, his task was to produce a corresponding sentence of type S. Thus, both discrimination and transformation were involved and presumably both affected the times required to perform the task. How much they affected the response times, of course, was strongly dependent on the relation between the two types of sentences. In most cases, however, it was not the physical similarity, as measured by identical stimulus or response elements, that seemed to matter; rather, the important features seemed to be of a semantic and syntactic character.

Is there some simple way to describe linguistic relations? Presumably, any notions we might have about structural relations, syntactic generalization, etc., will need to be made explicit in terms of some reasonable statement of rules for the language. For example, the linguistic structure of any sentence is essentially the set of rules that were applied in its construction; we might argue, therefore, that two entities are structurally related to the extent that the same rules were used to construct both. This definition suggests the possibility that, *ceteris paribus*, structurally related sentences, differing only with respect to one or two rules, might be easier to transform into one another than sentences differing with respect to several rules. By measuring the time required to perform such transformations, therefore, we might hope to derive an indicator of the extent of linguistic relatedness; the greater the number of rules that were different in the derivation of the two sentences, the greater the number of revisions that would be required to convert one into the other, and the longer the transformation would take. These were the general assumptions underlying the measurements reported here. Note, incidentally, that this argument is in some sense the inverse of the one we usually apply in measuring discrimination, where the greater the difference the shorter the discrimination time is expected to be.

Some linguists (Harris, 1952, 1957; Chomsky, 1955, 1957) have argued that at least two types of syntactic relation should be distinguished, one involving differences in syntactic rules for forming sentences, the other involving differences in rules of transformation. That is, one pair of sentences may be said to be syntactically related because many of the same rules were involved in their formation, whereas another pair are related because they were derived by applying different transformational rules to the same underlying linguistic structure. Although our choice of the particular syntactic aspects to be studied in the present experiments was influenced by this distinction between rules of formation and rules of transformation, the results of these experiments in no sense comprise a critique or evaluation of the distinction. Inasmuch as the distinction is one of theoretical convenience in simplifying a linguist's characterization of syntax, it is not obvious that any psychological evidence concerning the cognitive operations of the language user could be relevant to the question of the utility or validity of such a distinction in descriptive linguistics. A description of a language and a description of actual performances by language users must be kept distinct, much as number theory and practical methods of computation are kept distinct in mathematics, and for similar reasons. Of course, a language user's performance reflects his tacit knowledge of the language, but the actual procedures whereby he attempts to comply with the rules that he knows are little understood. Indeed, the psychological study of these processes has scarcely begun (see Miller and Chomsky, 1963).

### PRELIMINARY EXPERIMENTS

Our first attempts to devise a chronometric method to study syntactic relations, which were carried out in collaboration with D. Slobin, involved a form of pencil-and-paper test that could be administered to groups of people simultaneously; a preliminary account of this work is contained in a paper by Miller (1962).

In the first study, we took the set of eighteen three-word sentences that can be formed by using *He*, *She* or *It* as the first word, *hit* or *liked* as the second word, and *him*, *her*, or *it* as the third word: *He hit him*, *He hit her*, *He hit it*, *He liked him*, etc. In addition, we used the corresponding sets of sentences that can be produced from this set by passive, negative, and passive-negative transformations. Thus, for example, *She liked it* in the set of basic (active affirmative) sentences corresponded to *It was liked by her* in the set of passive sentences, to *She didn't like it* in the set of negatives, and to *It wasn't liked by her* in the set of passive-negatives. Because the same words occurred repeatedly in these sentences, it was assumed that, in order to make a correct match, the subject would have to operate on the basis of their syntactic features and not on the basis of particular words.

A test was constructed by taking two of these four sets of 18 sentences and arranging them in two columns vertically on a page. Half of the 18 sentence pairs were arranged with one type of sentence in the left column, the other half with the other type in the left column; in the right column were the corresponding members in scrambled order. Before the page was presented to the subject he was asked to study a sample pair of sentences that illustrated the kind of similarity to be tested, and he prepared himself to search for such similarities between the two columns. When the signal was given to start, he turned the page in his test booklet and began with the first sentence at the top of the left column; he decided what the corresponding sentence should be, searched and found this sentence in the right column, and then placed the number of the corresponding sentence to the left of the original sentence in the left column. He continued this way down the left-hand column until, at the end of one minute, he was instructed to stop. He then turned the page and studied the instructions for the next test. His score on the test was the number of sentences correctly matched; from that it was possible to estimate the average time he required per sentence.

With four sets of sentences there are six comparisons to be tested: (1) active affirmative sentences (AA) against passive affirmative sentences (PA); (2) AA against active negative (AN); (3) AA against passive negative (PN); (4) PA against AN; (5) PA against PN; and (6) AN against PN. In addition, four control tests were included that involved no syntactic differences; the subject simply read the sentence and then searched for that identical sentence in the right-hand column: (7) AA against AA; (8) AN against AN; (9) PA against PA; and (10) PN against PN. These control tests involved all the operations of reading, remembering, searching, and writing that were involved in the tests with different sets of sentences—everything except the operations that were required in order to cope with the syntactic differences between the sets of sentences. In one important respect, these controls proved inappropriate: should the times estimated when sentences of type *S* are tested with sentences of type *S′* be compared with the control data for type *S* or for type *S′*? We decided to use the average control times for *S* and *S′*, but we do not regard this solution as completely satisfactory. It was assumed, however, that the differences in times per sentence between each of the six experimental tests and the appropriate combinations of the four control tests would provide a measure of the additional time required to cope with these syntactic differences.

Sixty subjects took this battery of 10 tests, plus one additional practice test at the beginning to acquaint them with the test procedures. The results indicated that 0·18 sec. additional time per sentence was required when matching AA and AN sentences, and that 0·82 sec. additional was required for matching PA and PN sentences; the average time required to cope with an affirmative-negative difference, therefore, was 0·50 sec. Similarly, the active-passive difference required 3·70 sec. additional time on the average; 2·13 sec. for matching AA and PA, and 5·26 sec. for AN and PN. When both differences were involved, the average additional time required was 3·56 sec.; 2·25 sec. for matching AA and PN, and 4·86 sec. for matching PA and AN.

Although these results were encouraging, in that they showed there was something measurable here, there were several shortcomings in this study: (1) The use of pronouns was especially unfortunate, since it unnecessarily complicated the passive transformations; for example, *He liked her* can be permuted to form *Her was liked by he* according to the usual transformational rule for the passive, but then *Her* must be further changed to *She* and *he* to *him* before the result is grammatical. These additional modifications probably led to a considerable overestimation of the times normally required to deal with an active-passive difference. (2) The data showed great variability; some subjects completed few of the matches correctly, and so contributed very long response times to the group averages. Moreover (3) the control conditions used were not completely appropriate and a somewhat dubious average of two control tests was necessary in order to provide a base line against which to estimate the additional times required in the experimental conditions when two different sets were matched.

There is something obviously artificial about using written symbols to study linguistic rules, so Slobin (1964) undertook a pilot study with sentences presented acoustically and responses made vocally. The two directions of transformation (e.g. AA changed to AN and also AN changed to AA) were measured separately for each comparison; the subject was not required to discriminate between the two types of sentences in order to decide which direction of transformation would be required. The time between the end of the presented sentence and the beginning of the subject's transformed response was measured with a stopwatch for eight subjects. The additional time measured in this manner was considerably shorter: 0·32 sec. of additional time was required to deal with the affirmative-negative difference, 0·82 sec. with the active-passive difference, and 1·03 sec. when both differences were involved. No statistically significant differences were observed as a function of the direction of the transformations. In order to obtain more accurate measurements of response times, several instrumentation problems were involved, so this auditory-vocal procedure was not pursued further at that time.

Instead, a revised and improved version of the pencil-and-paper procedure was explored next. Eighteen sentences were formed by taking *Jane*, *Joe*, or *John* as the subject, *liked* or *warned* as the predicate, and *the small boy*, *the old woman*, or *the young man* as the object. The procedure was much the same as before, except that the sentences to be transformed were presented on the bottom half of the page, and the sentences to be searched were in the top half, which should have simplified the mechanics of searching and writing. Also, the control tests were designed somewhat differently; there were six control tests, one for each of the experimental tests, and each control involved exactly the same search list as the corresponding experimental test, so that all conditions in the control and experimental tests were precisely the same except that on the control tests there were no syntactic differences between the sentence presented and the sentence to be found.

Again, 60 subjects were run, and the average number of matches completed in one minute was used to estimate the time taken per sentence. The additional time required to cope with affirmative-negative differences was 1·40 sec. on the average; 1·13 sec. when matching AA and AN sentences, and 1·66 sec. when matching PA and PN. For the active-passive difference, the additional time required was 1·65 sec. on the average; 1·43 sec. for matching AA and PA, and 1·87 sec. for matching AN and PN. And, finally, when both differences were involved, an additional 3·12 sec. was required on the average; 2·74 sec. for matching AA and PN sentences, and 3·50 sec. for matching PA and AN sentences.

The additional times required for the active-passive difference were considerably reduced from those obtained with our first pencil-and-paper tests, as expected, although the affirmative-negative difference seemed to take longer. It is especially interesting to note that when the 1·40 sec. required to deal with affirmative-negative differences is added to the 1·65 sec. required to deal with active-passive differences, the result is 3·05 sec., reasonably close to the 3·12 sec. of additional time that was measured when both of these differences were involved. These figures suggest the very strong hypothesis that the time required for dealing with these syntactic differences is linearly additive.

Affirmative-negative and active-passive differences are, according to some conceptions of English grammar, created by transformational rules, i.e. the passive can be derived from the active (usually without any semantic change), or the negative from the affirmative (with a particularly significant semantic change), by applying a special type of transformational rule to the abstract structures underlying these sentences. (The status of negation—formational or transformational—is debatable, however. Even among linguists who accept transformational rules in principle there are many who would regard negation as a preverb which, like an auxiliary verb, is introduced by a rule of formation rather than by a subsequent application of a rule of transformation.)

Other types of syntactic similarity are not considered to have this transformational character; for example, *John liked the small boy* and *John had been liking the small boy* are not related by a transformational syntactic rule, but are generated by different rules of formation (and entail a clear semantic difference). In order to see whether this difference in linguistic status corresponds to any measurable difference in performance, therefore, we conducted still another pencil-and-paper study.

From the 18 sentences constructed from *John, Jane*, or *Joe liked* or *warned the small boy, the old woman*, or *the young man* we constructed three other sets of sentences differing in the verb construction. Let us refer to the reference set of sentences as simply p, for past tense. Another set was derived from p by adding the auxilliary verb *be* (bp); another by adding the auxiliary verb *had* (hp); and a third by adding both *had* and *be* (hbp). For example, *Jane warned the old woman* in set p corresponded to *Jane was warning the old woman* in set bp, to *Jane had warned the old woman* in set hp, and to *Jane had been warning the old woman* in set hbp. Twelve tests, six experimental and six control, were constructed as before, and, as before, 60 subjects were tested. In all respects except the nature of the syntactic differences among the sets of sentences, the study was conducted precisely as before.

The results showed that on the average 0·86 sec. of additional time is required when the sentences to be matched differ by having *be* constructions in them; 0·87 sec. extra was required when p and bp sentences were matched, and 0·85 sec. when hp and hbp were matched. When the difference involved the use of *had* in the verb, the average additional time required was 0·97 sec.; 0·83 sec. when matching p and hp sentences, and 1·11 sec. when matching bp and hbp. When the sentences to be matched differ in both respects, the additional time required was 1·43 sec. on the average; 1·68 sec. when matching p and hbp, and 1·18 sec. when matching bp and hp.

These differences were dealt with slightly more rapidly than were the affirmative-negative and active-passive differences. More significant, however, is the fact that the additional time required to deal with two syntactic differences does not seem to be the sum of the extra time needed for each alone; except for the value of 1·68 sec. observed for matching p and hbp sentences, all of the differences seem relatively uniform in magnitude; even this exception may well be attributable to errors of measurement. An analysis of variance using the experimental minus the control score for each individual on each test (item scores were more nearly normally distributed than time scores) indicated that there were no significant differences between sentence types in this experiment, whereas the value of $F$ obtained in the experiment with passive and negative sentences had been highly significant. These results, if confirmed in further studies, might be used to argue that, regardless of the status of the formational-transformational distinction in descriptive linguistics, these two types of syntactic relations are dealt with differently by the people who use the language.

One serious difficulty with these pencil-and-paper tests was the variability in the data from different subjects. Thus it was necessary to test at least 60 subjects in order to obtain stable averages; much larger samples would have been required for a truly definitive study. A major source of this variability, of course, was the search task. Even under the best of conditions, the length of time required to complete a search will be uncertain. Moreover, this source of variance is of no intrinsic interest, since the additional time required to deal with syntactic differences was not (in theory, at least) dependent upon the search in any way. We were extremely grateful to Donald A. Norman, therefore, when he suggested that the sentences to be transformed should be presented one at a time, and provided us with equipment suitable for measuring the reading times and the search times separately.

## EXPERIMENTAL EQUIPMENT AND PROCEDURE

The device used to present the sentences and to measure the subject's reaction times can best be described as a two-field tachistoscope, where the duration of exposure of each field was under the control of the subject. The top quarter of the total field displayed the sentence that was to be transformed; when he was ready, the subject pressed a hand-held switch which illuminated this sentence. When he had read it, performed whatever operations on it were required, and was ready to search for the matching sentence, he pressed the switch again; this turned out the light in the presentation field and turned it on in the bottom three-quarters of the field, where the set of 18 sentences to be searched through were presented. When he found the corresponding sentence, he pressed the switch again, announced its number on the list, and waited for the next presentation.

The sentences were typewritten using carbon paper in such a way that they were printed on the reverse side of the paper; room illumination was low enough that the printing could not be seen through the paper until the illumination was turned on behind the translucent plastic surface against which the paper was pressed by a transparent glass cover. Although the printing was slightly fuzzy and the contrast relatively low, the sentences were all quite legible to the subjects, whose eyes were about 12 to 15 in. from the display surface.

The switch controlled not only the illumination of the sentences, but also two clocks visible only to the experimenter. One clock recorded the time the subject spent studying the presented sentence, the other recorded the time spent searching. Between trials the experimenter recorded these two times (to the nearest 0·01 sec.), recorded the subject's response, and changed the sentences for the next trial according to a pre-arranged random order. Ten sentences in succession were tested with each syntactic difference that was being studied; the same instructions and search list were used throughout this sequence of ten trials; only the presentation sentence was changed on every trial.

At the beginning of the experiment the following instructions were read:

"This is an experiment to test your ability to perform some sentence transformations. It is important that you work as accurately and as quickly as you can.

"Before each of the subtests I will place a set of sentences in the bottom section, like this, and then place a single sentence in the top section, like this. You can see the sentences only when a light is turned on behind them. This you can do by pressing this button with your thumb. The first time you press it the single top sentence is lit, like this. When you press it again, the top light goes off and the bottom one comes on, so you can see the set of sentences, like this. Your job is to match the single sentence at the top to one of the bottom sentences in the manner in which I tell you. For example, I will give you a list of sentences and tell you to find the sentence that means the same thing as the sentence at the top, but in a different form. For example, *He caught her* means the same thing as *She was caught by him.* Before each subtest I shall give you an example of the form the sentences will take, and will give you several examples so that you will know which transformations you are to make. You will have to transform sentences in both directions in these subtests. In some of the subtests you will be asked to find the identical sentence in the bottom set, but each time I will tell you what you are to do.

"After I have placed the first single sentence in the top section, you are to press the button, read the sentence, make the transformation, press the button again, search quickly through the bottom set until you find the correct sentence (the one that means the same thing, but is in the required form). The important thing is to make the required transformation before you press the button to start searching through the bottom set. You are to search as quickly as you can until you find the correct sentence. When you find the correct sentence, press the button again and tell me what sentence it is. Then wait until I put in another single sentence. Remember that you must be accurate as well as fast, for I am measuring the amount of time that elapses between button presses and also recording your answer."

The first set of 10 sentences was a practice trial and the data were discarded. In each experiment ten subjects were run individually; they were all undergraduate students at Radcliffe College and Harvard University whose first language was English. They were paid $1.50 for the experimental session, which generally lasted about one hour.

The test materials were precisely the same as those used in preliminary experiments with pencil-and-paper tests: eighteen sentences were formed by taking *John*, *Jane*, or *Joe* as the first word, *liked* or *warned* as the second, and *the small boy*, *the old woman*, or *the young man* as the final phrase. In Experiment 1 three other sets of eighteen corresponding sentences were formed by taking the negative, passive, and passive-negative forms of each sentence in the original set. In Experiment 2, three other sets of eighteen sentences were formed from these by adding to the verb *be*, *had*, or both.

From these sets of sentences search lists were formed. For example, if the difference between active-negative (AN) and passive-negative (PN) sentences was to be studied, the AN member of half the pairs and the PN member of the other half of the pairs would appear on the search list. For the experimental condition, the corresponding members of these pairs would be presented as single sentences to be transformed; in the control conditions, the same search list was used, but the sentences presented singly were identical with the sentences on the search list and no transformation was required. (In this control, different (i.e. transformed) sentences were presented singly and the same search list was used as in the experimental condition. A good case can also be made, however, that the control condition should present exactly the same sentences singly and use a transformed search list. Our choice was dictated by our desire to obtain data comparable to those obtained in the pencil-and-paper tests, but the methodological point involved deserves further investigation.)

In each experiment 13 subtests were given: one initial test for practice, six experimental tests comparing all possible pairs of sentence types, and six control tests involving no syntactic differences between the sentence presented and the sentence to be found. The order of the experimental and control tests was scrambled for each subject, but care was taken not to run an experimental and its own control test in immediate succession.

## RESULTS

The data of major interest are those for the presentation times required to read and transform sentences prior to beginning the search task, but let us glance first at the search times. Search averaged about 4·5 sec. per sentence, with considerable variance; differences between the times required to search under experimental conditions and under the corresponding control conditions ranged from −0·22 to 1·69 sec., with an average of 0·67 sec. for the first experiment and 0·31 sec. for the second. No intelligible pattern could be discerned in these differences as a function of the syntactic variables involved. Perhaps some pattern exists, but search times are variable and a sample of more than ten subjects would be required to reveal it. We discarded the search data, however, not only for statistical reasons, but also because we had no theoretical justification for believing the search times are relevant to the syntactic variables that are the central object of this research.

The average presentation times that the subjects required in Experiment 1 are given in Table I, along with the differences between the experimental and the control conditions that we have taken as a measure of the additional time involved in dealing with the syntactic differences among the sentences. All these differences (except for PA → PN) are significantly greater than zero at the 0·05 level. The matrix of differences is reasonably symmetrical, which confirms Slobin's observation that the direction of the transformation is not an important factor influencing the results. If we average the two directions, therefore, we obtain the following figures. An additional 0·41 sec. was required to deal with the affirmative-negative difference (0·40 sec. when matching AA and AN sentences, and 0·42 sec. when matching PA and PN). An additional 0·91 sec. was required to deal with the active-passive difference (0·81 sec. when matching AA and PA sentences, and 1·01 sec. when matching AN and PN). When both differences are involved, the additional time required is 1·53 sec. (1·24 sec. when matching AA and PN sentences, and 1·82 sec.

## TABLE I

### AVERAGE PRESENTATION TIMES IN SEC. FOR EXPERIMENT 1

Rows are labelled according to syntax of sentences presented; columns are labelled according to syntax into which it had to be transformed (experimental condition) or which also occurred on the same subtest (control condition).

| Experimental: | | AA | AN | PA | PN |
|---|---|---|---|---|---|
| | AA | — | 1·81 | 2·20 | 2·49 |
| | AN | 2·08 | — | 3·60 | 2·73 |
| | PA | 2·51 | 3·51 | — | 2·20 |
| | PN | 2·92 | 2·72 | 2·37 | — |
| Control: | | | | | |
| | AA | — | 1·42 | 1·46 | 1·35 |
| | AN | 1·68 | — | 1·80 | 1·70 |
| | PA | 1·63 | 1·68 | — | 1·80 |
| | PN | 1·58 | 1·73 | 1·93 | — |
| Differences: | | | | | |
| | AA | — | 0·39 | 0·74 | 1·14 |
| | AN | 0·40 | — | 1·80 | 1·03 |
| | PA | 0·88 | 1·83 | — | 0·40 |
| | PN | 1·34 | 0·99 | 0·44 | — |

when matching PA and AN). If we add the 0·42 sec. additional time required for the affirmative-negative difference to the 0·91 sec. additional time required for the active-passive difference, we obtain 1·33 sec., somewhat less than the 1·53 sec. actually observed when both differences were involved.

Note that we have calculated here differences between: (*a*) the time required to read *S* and transform it into *S*′, and (*b*) the time required to read *S* (when *S*′ might have occurred). It is also possible, of course, to calculate the differences between (*a*) and (*c*) the time required to read *S*′ (when *S* might have occurred), by subtracting the transpose of the control matrix from the experimental matrix. When this is done, the matrix of differences is no longer symmetrical; it always takes more time, according to this calculation, to transform a longer sentence into a shorter one than vice versa. We assume that this asymmetry is due to the fact that differences in reading times as a function of sentence length are not properly controlled in the (*a*–*c*) comparison, and for that reason we have instead presented the (*a*–*b*) comparison in these tables.

Note also that we might have used for our controls the average times required to read *S*, where the average is computed over all the three alternative contexts in which *S* was tested. The differences in reading times as a function of context in the control tests were not statistically significant, so such averaging for each row would be justifiable. When this is done, however, the matrix of differences remains symmetrical and the interpretation of the results is unaffected.

The average presentation times in Experiment 2 are given in Table II, again with differences indicating the additional time required to deal with the syntactic variables. All these differences (except for hp→p) are significant at the 0·05 level. Once again we see that the matrix of differences is reasonably symmetrical—almost uniform, in fact, except for the p–hp entries. Averaging as before, we obtain the following figures. The additional time required to deal with the occurrence of *had* in the verb was 0·37 sec.; 0·19 sec. when matching p and hp sentences, and 0·54 when matching bp and hbp. An additional 0·63 sec. was required to deal with the occurrence of *be*, both when matching p and bp and also when matching hp and hbp sentences. When both *had* and *be* were different, an additional 0·60 sec. was required; 0·53 sec. when matching p and hbp sentences, and 0·67 sec. when matching hp and bp. There is little hint of additivity in these data.

As in the preliminary studies, error rates were around four per cent; errors did not seem to be more frequent on the more difficult comparisons, but there were too few errors to establish any clear trend. Trials on which errors occurred were omitted in computing the averages presented in Tables I and II.

TABLE II

Average Presentation Times in Sec. for Experiment 2

Rows are labelled according to the verb construction of the sentence presented; columns according to the verb into which it had to be changed (experimental condition) or which also occurred on the same subtest (control condition).

| Experimental: | | p | hp | bp | hbp |
|---|---|---|---|---|---|
| | p | — | 1·58 | 1·88 | 1·82 |
| | hp | 1·72 | — | 2·31 | 2·06 |
| | bp | 1·94 | 2·14 | — | 2·11 |
| | hbp | 2·26 | 2·37 | 2·22 | — |
| Control: | | | | | |
| | p | — | 1·33 | 1·26 | 1·40 |
| | hp | 1·60 | — | 1·64 | 1·45 |
| | bp | 1·29 | 1·47 | — | 1·52 |
| | hbp | 1·63 | 1·73 | 1·72 | — |
| Differences: | | | | | |
| | p | — | 0·25 | 0·62 | 0·42 |
| | hp | 0·12 | — | 0·67 | 0·61 |
| | bp | 0·65 | 0·67 | — | 0·59 |
| | hbp | 0·63 | 0·64 | 0·50 | — |

## Discussion

Concerning this "tachistoscopic" method of investigating syntactic similarities, it should be remarked that more stable data were obtained with 10 subjects than were obtained with 60 using pencil-and-paper tests. Moreover, the visual display is more flexible and could be used to study a variety of other similarities that might be difficult to formulate as pencil-and-paper tests.

The general pattern of the data is much the same as that observed in the preliminary tests. In Experiment 1 the passive construction took more additional time than the negative, and both together required about as long as their sum separately. We cannot say whether this additivity would appear with other transformations—certainly it does not appear in Experiment 2 when only rules of formation are involved—but where it is observed, it strongly suggests that some kind of serial processing is taking place. Also, as before, we note in Experiment 1 that matches involving active affirmative sentences were always shorter (required less additional time) than the corresponding matches that did not involve active affirmative sentences: AA–PA was less than AN–PN; AA–AN less than PA–PN; and AA–PN less than PA–AN. This consistent bias suggests that active affirmative sentences may in some sense be psychologically primary and easier to cope with.

In the results of Experiment 2 we note that all the conditions required approximately 0·6 sec. additional time, except for the p–hp comparison, which required only 0·2 sec. This p–hp comparison was the only one of the six that involved only a simple insertion or deletion of a single word, e.g. *John liked the small boy* vs. *John HAD liked the small boy*. In all the other cases tested, the differences were more complicated and necessitated substitutions, e.g. in *John was liking the small boy* vs. *John had been liking the small boy*, the subject must interchange *was* and *had been*.

It would seem, therefore, that substitution is more complicated (takes longer) than insertion or deletion of a single word.

Under the conditions of Experiment 1, we can conclude that the processing time is about one second longer for passive than for active sentences, and about 0·4 sec. longer for negative than for affirmative. With this chronometric method, therefore, it might seem that the passive transformation takes a second and the negative takes 0·4 sec. to perform. If we were to leave the matter at this point, however, with the implication that these reaction times are characteristic of passive and negative constructions wherever they occur, the description would be dangerously over-simplified.

The difficulty can be illustrated by reference to related studies. Wason (1959, 1961, 1963) has reported that the time required to evaluate the truth of a sentence is more affected by its syntax (affirmative or negative) than by its truth value (true or false). If negation has a delaying effect in Wason's experiments, then, on the basis of the results reported here, passive constructions might be expected to cause even greater delays. However, McMahon (1963) has measured the times required to judge whether written sentences were true or false; in a series of three experiments McMahon found that passive constructions took only 0·1 to 0·2 sec. longer to judge than did corresponding active constructions of equal length, whereas negative sentences required from 0·3 to 0·5 sec. longer than affirmative. When McMahon's results are compared with those reported here, about the same amount of additional time is required to deal with negative sentences and the pattern of linear additivity is roughly maintained for PN sentences, but the predicted severe difficulty of the passive sentences was simply not observed by McMahon. Measured in this manner, therefore, active and passive sentences seem more closely related than affirmative and negative.

This result has been confirmed by Slobin (1964), who tested adults and children from age 6 to 12 years by presenting pictures and speaking true or false sentences about the pictures. At all ages the negative sentences took longer to evaluate than did the passive. Moreover, Slobin noted that when there is little chance of confusing the actor with the object acted upon, passive could be evaluated just as rapidly as active sentences. That is, for "reversible" sentences (e.g. *John hit Jane* or *Jane hit John*), adults required 0·15 sec. longer to evaluate passive than active forms, confirming McMahon's results which were also for reversible sentences. But for "non-reversible" sentences (e.g. *John ate lunch*, but not *Lunch ate John*), the passives did not take any longer than the active versions of the same sentences. Slobin suggests that the difficulty in understanding reversible passive sentences is probably attributable to the necessity for remembering which noun phrase is the actor, a difficulty that is solved semantically in the "non-reversible" sentences.

Taken together, these two sets of experimental observations strongly suggest that the additional times were spent largely in semantic rather than syntactic operations. Obviously, only semantic operations were involved in the verb constructions of Experiment 2 (except for the apparently artificial p–hp comparison, as noted above). In Experiment 1 the negative transformation is clearly semantic in effect and it seems to take about 0·4 sec. in both the matching and the evaluating tasks; in fact, the negative need not be considered a transformation at all, but can be regarded linguistically as a preverb—as part of the verb construction, just as in the case of the auxiliary verbs *be* and *have*. In an evaluation task the passive, which does not involve any semantic change, does not take any additional time unless there is some potential source of semantic confusion (e.g. the possible reversal of subject and object). In sum, the only main effect we cannot attribute to a semantic operation is that observed for the active-passive difference in the matching task. But even this difference is questionable. Is this extra time spent in performing a normal syntactic transformation? Or is this relatively long additional interval devoted to some artificial, non-syntactic, non-semantic permutation invented by the subjects just to satisfy the somewhat unusual requirements set by this particular experimental task? Surely the latter interpretation is no less compelling than the former.

The only clear evidence left that syntax contributed to the measured reaction times would seem to be the fact that matches involving active-affirmative sentences always required less additional time than did the corresponding matches not involving them. This result might be expected if every sentence interpreted (or produced) required a complete syntactic analysis (or synthesis), and if the active-affirmative sentences were syntactically simpler and so required less additional time for their analysis (and synthesis). This is, in essence, what a transformational theory says: passive, negative, and passive-negative sentences contain all the same syntactic rules as do active-affirmative sentences, plus one or two more which increase their complexity and require a little more time for interpretation (or production). If we base our estimates of the times required for syntactic transformations on this argument, the results are of the order of 0·1 sec. or less per syntactic transformation, and the observed bias is reasonably well accounted for. It should be well noted, however, that these syntactic operations are a normal and regular part of understanding and producing sentences and are not distinctively involved in the present experimental task of converting one kind of sentence into another.

The general impression we have formed of how people perform syntactic operations in normal usage is that these operations are quite automatic and involuntary, as are most of the information processes that psychologists have postulated to occur in perception or in motor skills. Any deliberate and explicit control that a person can exercise over these linguistic operations is principally, and perhaps exclusively, at the semantic level.

This research was supported in part by grants to Harvard University, Center for Cognitive Studies, from the Carnegie Corporation of New York, the National Science Foundation (NSF-G-16486) and the National Institutes of Health (MH-05120). We are grateful to Drs. D. I. Slobin and D. A. Norman for assistance in carrying out various phases of the research and to R. W. Brown, J. S. Bruner, J. B. Carroll, Z. S. Harris, J. Marshall, L. E. McMahon, A. G. Oettinger, F. Smith, P. C. Wason, and U. Weinreich for valuable criticism of an earlier informal report on this work. The paper was written while the senior author was visiting the Institute of Experimental Psychology, Oxford University, as a Fulbright Research Scholar.

## REFERENCES

CHOMSKY, N. (1955). *Logical Structure of Linguistic Theory*. Microfilm, Massachusetts Institute of Technology Library.

CHOMSKY, N. (1957). *Syntactic Structures*. The Hague.

HARRIS, Z. S. (1952). Discourse analysis. *Language*, **28**, 474–94.

HARRIS, Z. S. (1957). Co-occurrence and transformation in linguistic structure. *Language*, **33**, 283–340.

McMAHON, L. E. (1963). *Grammatical Analysis as Part of Understanding a Sentence*. Unpublished doctoral dissertation, Harvard University.

MILLER, G. A. (1962). Some psychological studies of grammar. *Amer. Psychologist*, **17**, 748–62.

MILLER, G. A., and CHOMSKY, N. (1963). Finitary models of language users. In LUCE, R. D., BUSH, R. R. and GALANTER, E. (Eds.), *Handbook of Mathematical Psychology*, v. 2. New York.

SLOBIN, D. I. (1964). *Grammatical Transformations in Childhood and Adulthood*. Unpublished doctoral dissertation, Harvard University.

WASON, P. C. (1959). The processing of positive and negative information. *Quart. J. exp. Psychol.*, **11**, 92–107.

WASON, P. C. (1961). Response to affirmative and negative binary statements. *Brit. J. Psychol.*, **52**, 133–42.

WASON, P. C., and JONES, S. (1963). Negatives: Denotation and connotation. *Brit. J. Psychol.*, **54**, 299–307.

*Manuscript received* 10*th June*, 1964

# STIMULUS SET AND RESPONSE SET: THE ALTERNATION OF ATTENTION

BY

D. E. BROADBENT and MARGARET GREGORY

*From the Applied Psychology Research Unit, Medical Research Council, Cambridge*

Recently Gray and Wedderburn showed that if a meaningful phrase of three words was presented together with three digits, in such a way that each ear received some items of each of the two types of material, it was no harder to recall the items grouped by type than it was to recall them grouped by ear. This finding is repeated and confirmed in several forms, culminating in the use of three letters of the alphabet and three digits as the six items presented. It is also shown, however, that even when all material is presented to one ear, it is harder to recall a list made up of alternate items of two classes than it is to recall the same items arranged as two successive sub-lists. Thus Gray and Wedderburn's result does not appear to reveal a situation in which alternation of attention between the ears is especially easy, but rather one in which continued attention to one ear is especially difficult because it requires alternation between classes of item.

A further series of experiments showed that a reduction in the presentation rate of stimuli produced a much greater improvement in performance when the items were of two alternated classes than when the classes were left separate. Equally, a slow presentation rate is more helpful when alternation between ears is required than when each ear is to be dealt with separately. These results support the idea that attention takes time to shift; but require a separate kind of attention, to a class of item rather than to a source of stimulation.

## INTRODUCTION

When two lists, each of three digits, are presented simultaneously to separate ears at speeds greater than one digit per ear per 1½ sec. it has been found (Broadbent, 1954; Moray, 1960) that recall in successive order (i.e. all stimuli on one ear followed by all those on the other) is superior to recall in alternating order (L R L R L R). Broadbent has interpreted these findings in terms of his filter theory. On that view a filter selects stimuli on the basis of some common physical characteristic such as arrival on the same sensory channel, and time is necessary to change from one channel to another. In the above situation when the interval between pairs is shorter than 1½ sec. there is not supposed to be sufficient time available for the filter to switch from one ear to the other and back again, and so successive recall is easier.

Gray and Wedderburn (1960) have questioned the generality of this interpretation. They suggested that since the material—single digits—offered no further cues for grouping, other than the ear of arrival, the superiority of successive recall does not necessarily reflect some built-in mechanism. Such grouping may be simply an optimal strategy for this situation, based on past success in classifying the input from one ear as a continuous message, since in everyday life the sources of meaningful messages do not change position repeatedly. In support of their interpretation they devised an experiment in which subjects were presented with phrases made up of three monosyllabic words, the words alternating between the two ears and each word being presented simultaneously with a digit in the opposite ear (e.g. Right: 1 Aunt 3; left: Dear 2 Jane). More completely correct lists were achieved when the order of recall grouped items by meaning, i.e. three words followed by three digits, than when the grouping was by ear of arrival. This superiority of grouping by meaning was more marked in subjects who had been informed that the three words made up a phrase.

In view of the interest of these results, the present authors have carried out a series of experiments to obtain more information about the conditions under which the effect appears.

## PRELIMINARY EXPERIMENTS

Two studies, each of twelve subjects were conducted to rule out artefacts which might possibly have entered into the procedure used by Gray and Wedderburn. Since in fact the results showed that the effect still appeared under the modified conditions, these experiments are not reported in detail. The differences from the earlier report were as follows. The subjects were required by instruction to respond in one order during one session and the other order in the other session, rather than leaving the recall order to the preference of the individual subject; and the vocabulary of words used in the meaningful phrases was confined to a few definite words. These were so chosen that any of the words occurring in the first position could be combined with any of the other words in the second or third position: thus hearing one word did not improve the listener's chance of guessing any of the other words. Nevertheless it was at least as easy to report the items grouped into digits and phrase as it was to report ear by ear.

In the second study, two phrases, rather than one phrase and three digits, were interlaced in the style of Gray and Wedderburn. Once again performance was, if anything, better when phrase by phrase recall was employed. Thus the effect does not seem to be due solely to the use of different vocabularies such as digits and words. However, two important new points arose from these preliminary experiments.

Firstly, it was clear that performance in the ear-by-ear condition was far worse than it was in the original experiment of Broadbent (1954), although the apparatus and subjects were similar and the vocabulary was a small and known one. (This low level of performance was also notable in Gray and Wedderburn's results, but as their experimental conditions might have differed in a number of ways from those of Broadbent no clear interpretation could be put on the fact.) It thus appeared that some feature of Gray and Wedderburn's procedure was *impairing* performance in the ear-by-ear case rather than making it easier in the phrase-by-phrase case.

Secondly, it was suggested to the authors (by D. R. J. Laming) that the use of phrases in this type of experiment leads to an ambiguity of interpretation. A meaningful phrase can be reformulated by the listener into a single unit, such as a visual image, in a way that three isolated words cannot. From the results reported by Miller (1956) this would mean that better performance would be expected even for items all arriving by the same sense-organ, if they were grouped by meaningful units rather than in a random sequence of words; Miller, it will be remembered, taught subjects to reformulate sequences of zeros and ones into decimal digits, with a great improvement in the memory span. Thus there should be an advantage in memorizing the phrases rather than the isolated words even in the single-channel case; the authors carried out a brief experiment to confirm that this is indeed so, but the point is fairly obvious once it has been made. This advantage would work against any difficulty of shifting attention in Gray and Wedderburn's experiment, and would make it impossible to tell whether such a difficulty exists or not. This factor would not, however, explain the poor performance found in the ear-by-ear condition as compared with Broadbent's original situation.

These two points suggested a new line of experimentation, in which classes of words rather than continuous meaningful phrases were interweaved in the Gray and Wedderburn manner.

## EXPERIMENT I: SHIFTING ATTENTION BETWEEN CLASSES OF WORDS

### *Method and general procedure*

A two-channel Ferrograph tape recorder was used, each channel being led to one ear of a split head-set. All subjects were Naval Ratings, and each worked under one condition on one day and another condition on the next day. The order of conditions was counterbalanced between subjects. Whatever the condition, each trial in a session consisted of six items being presented to the listener and the tape recorder then being stopped while he wrote down as many of the items as he could recall. The rate of presentation was two items per ear per sec. Ten such trials made up one day's session; before the session the condition being used was explained to the subject and, in Group I*a*, two practice trials were given and the responses corrected. The subjects used in Group I*a* had had no previous experience with this type of task; those in Group I*b* had served previously in the second of the preliminary studies.

*Group* I*a* (10 subjects). In this group three of the six items in each list were letters and three were digits. A letter presented to one ear was simultaneous with a digit to the other ear, and the type of item alternated between ear, e.g.

Right A 4 J
Left 1 F 9

The letters were chosen from the first 13 letters of the alphabet omitting B, G, and I: and the digits from random number tables. No letters or digits occurred twice in the same trial.

The same tape was used in both conditions but in one session the listeners were instructed to write down all items on one ear before any on the other ear. In the other condition the instructions were to recall all letters first and then all digits.

*Group* I*b* (12 subjects). In this case only the right-hand ear was stimulated. In one condition each trial consisted of three colour names followed by three digits. The colour names were chosen from the following: red, white, orange, green, black, blue, pink, yellow and grey. As before, selection was random subject to the constraint that no item occurred twice in the same trial and each colour name occurred in at least three trials. The instructions were to write down the items in the actual order of arrival, e.g. red white orange 496.

In the other condition the recording consisted of the same items in each trial, but arranged with colours and digits alternately, e.g. red 4, white 9, orange 6. As before subjects were to write down items in the order of arrival.

### *Results and discussion*

The results are shown in Table I. It will be seen that in Group I*a* recall ear-by-ear was only just superior to recall of letters and digits separately: the difference was quite insignificant, $p > 0.2$ both for sign test and by the more sensitive method of taking Session 2–Session 1 for each subject and computing a Mann-Whitney U between the two sub-groups. Thus once again Gray and Wedderburn's result is found, that it is no easier to stick to one ear than to alternate between ears. Once again both conditions give a very poor level of performance (about one-third of that attained in Experiment II or in Broadbent (1954) by subjects performing ear-by-ear recall of six digit lists).

In Group I*b*, alternating classes of item is clearly inferior to grouping the items by their type, $p < 0.004$ by sign test.* But alternation of classes of items is required of Group I*a* when they perform ear-by-ear recall. It is not surprising therefore that the latter type of recall is much worse then it is in similar situations involving only digits, and also that it shows no superiority over alternating between ears. There is of course a considerable difference between the general level of performance in

---

* Subsequent to this experiment, the authors discovered that similar findings had been obtained in unpublished work both by R. Conrad and by P. M. Rabbitt. Furthermore, since the paper was submitted, results similar to Experiment I (a) have been published by Ynetema and Trask, *J. verb. Learn verb. Behav.*, **2**, 65–74.

TABLE I

EFFICIENCY OF RECALL FOR MIXED LETTERS AND DIGITS WHEN PRESENTED DICHOTICALLY FOR RECALL IN DIFFERENT ORDERS, AND WHEN PRESENTED MONAURALLY IN DIFFERENT ORDERS

(*a*) DICHOTIC PRESENTATION—EXPERIMENT I(*a*)

PER CENT. CORRECT

| *Order of recall* | *Complete lists* | *Letter items* | *Digit items* | *Mean items* |
|---|---|---|---|---|
| Ear-by-ear . . | 20 | 58·3 | 75·3 | 66·8 |
| Letters/Digits . . | 18 | 57 | 66·6 | 61·8 |

(*b*) MONAURAL PRESENTATION—EXPERIMENT I(*b*)

PER CENT. COMPLETE LISTS CORRECT

| 3 *Colours*/3 *Digits* | *Colours and Digits Alternated* |
|---|---|
| 96 | 80 |

Group I*b* and that in Group I*a*: 80 per cent. and 20 per cent. respectively for the conditions requiring alternation of types of item. Dichotic presentation normally produces an impairment of this type, which was shown by Broadbent (1954) to occur even when the digits to be recalled were simultaneous with quite irrelevant digits to the opposite ear. Thus it may be ascribed either to cross-lateral masking or to distraction. The effect of alternating between classes of items is superimposed on that of dichotic presentation, and appears to be about the same size in dichotic and monaural cases. To make this comparison, we should transform the various percentages to probits in order to avoid the limiting effect of perfect performance. After such transformation, the difference between the two conditions of Group I*b* is of approximately the same size as that between the performance of Group I*a* and the level reached in Experiment II for homogeneous dichotic material.

One may conclude that the Gray and Wedderburn effect shows that it is difficult to alternate attention between two classes of material just as it is to alternate between ears. It does not show that alternation between ears is easy given additional cues.

## EXPERIMENT II: EFFECTS OF PRESENTATION RATE UPON ALTERNATION OF WORD CLASSES AND OF SENSORY CHANNELS

As has been said, Broadbent explained the difficulty of alternating between ears as due to a time taken to switch attention, and attempted to support this by demonstrating that recall in an order alternating between ears was easier when the presentation rate was slower. If a similar explanation is to be advanced for the difficulty of alternating classes of word such as numbers and letters, it ought similarly to be possible to show that alternation is easier at slower speeds. However, it has been well pointed out by Moray (1960) that the original experiment of Broadbent (1954) was at fault in that the effects of presentation rate were only considered using an order of recall alternating between ears. Moray demonstrated that even recall ear-by-ear improves when presentation rate is low and that therefore the effect of rate might be a general feature of memory rather than one specific to alternation between ears. It needs to be shown therefore that the effects of slowing presentation rate are *more* beneficial when alternated classes of words are being heard than when groups of words of the same class are presented together. The following experiment was therefore designed to examine this question.

In addition, the opportunity was taken of reopening the original question of switching time between the ears. Just as Broadbent (1954) used only alternating order of recall, so Moray (1960) only used ear-by-ear recall. The question remained open whether a slow rate of presentation was *more* advantageous for one order of recall or for the other.

### *Method and procedure*

Using the same twin channel Ferrograph recorder each channel being connected to one of the phones of a split headset, two types of material were presented.

(1) *Dichotic lists.* This was the same material used by Broadbent (1954) to show that recall when alternating between channels is improved by slower presentation rates. It consisted of 24 lists each of three pairs of digits, each pair being presented simultaneously and dichotically. The rate of arrival of the pairs was varied systematically: lists 1–3 had one pair every 2 sec., lists 4–6 one pair every 1½ sec., lists 7–9 one pair every 1 sec., lists 10–12 one pair every ½ sec. In the second half of the tape the order was reversed, each triplet getting slower until the original speed of one pair every 2 sec. was reached. Thus there were six lists at each of the four speeds; three in the first half of the tape and three in the second. The subject was warned before each change of speed.

(2) *Monaural mixed letter and digit lists.* Two sets of 24 lists were prepared, each list having three letters and three digits and being presented to the left ear from one channel of the recorder. The letters were selected from the first 12 letters of the alphabet omitting D and I. In one set of lists the letters and digits were arranged successively, i.e. three letters followed by three digits, and in the other set they were arranged alternately. The speed of presentation varied systematically in the way described above except that the time intervals given were between individual items and not pairs of items.

Fifteen subjects were tested, all Naval Ratings. Each subject received the 24 dichotic lists: seven subjects were instructed to recall in alternating order, that is, both members of the first pair followed by both members of the second and then both members of the third. The other eight subjects were told to recall all that they heard on one ear followed by all that they had heard on the other. (The first two subjects in each of the above groups were in fact tested on two successive days, having the other condition on their second day, but transfer effects were found to be too serious to continue this procedure.)

Ten of the same subjects were presented with the monaural lists of letters and digits immediately after they had received the dichotic ones. Five of the subjects, who had been required to recall the dichotic lists in alternating order, were given the alternating presentation of letters and digits. The other five subjects, who had just been recalling ear-by-ear, were given the successive presentation of letters and digits. Before each set of 24 lists the subject was given one practice list at the slow speed.

As before, subjects wrote down their answers at the end of each list, the recorder being switched off as soon as the subject started writing. They were not allowed to fill in gaps or make corrections.

For reasons which will be given in the results section, a further group of six subjects, all members of the scientific staff of the A.P.R.U., were presented with the dichotic all-digit lists. The subjects were tested on two successive days: half of them recalled in the ear-by-ear order on their first day and the alternating order on their second day. The other three subjects were tested in the opposite order. There appeared to be little transfer in this case, possibly because all-over performance was much better. The subjects gave their answers verbally instead of writing them down.

### *Results*

#### (1) *Dichotic all digit lists*

(*a*) *Naval Rating subjects.* The mean performance for each group is shown in Table II. It is obvious that when alternating between the ears the subjects do better at the slowest speed whereas when recalling ear-by-ear they do better at the fastest speed. Two methods of scoring were used: the number of lists which were wholly correct, and the number of individual digits correct. The latter score is regarded by the present authors as unsatisfactory because it does not control the subject's strategy, but it is the one used by Moray.

TABLE II

PERFORMANCE UNDER ALTERNATING AND SUCCESSIVE CONDITIONS FOR MATERIAL DIVIDED BETWEEN THE TWO EARS AND FOR MATERIAL DIVIDED BETWEEN TWO CLASSES OF ITEM

(*a*) PER CENT. LISTS CORRECT

| | *Alternating* | | | | *Successive* | | | |
|---|---|---|---|---|---|---|---|---|
| | *Slow* | | | *Fast* | *Slow* | | | *Fast* |
| *Items per ear per sec.* | 1 *per* 2 *sec.* | 1 *per* 1½ *sec.* | 1 *per* 1 *sec.* | 1 *per* ½ *sec.* | 1 *per* 2 *sec.* | 1 *per* 1½ *sec.* | 1 *per* 1 *sec.* | 1 *per* ½ *sec.* |
| Dichotic lists Naval Ratings .. .. | 33·3 | 23·8 | 23·8 | 14·3 | 35·4 | 27·1 | 37·5 | 58·3 |
| Dichotic lists Lab. subs. .. .. | 91·7 | 77·8 | 47·2 | 33·3 | 69·4 | 66·7 | 58·3 | 66·7 |
| Monaural mixed class lists .. .. | 93·3 | 93·3 | 90·0 | 70·0 | 83·3 | 86·7 | 83·3 | 100 |

(*b*) PER CENT. ITEMS CORRECT

| | *Alternating* | | | | *Successive* | | | |
|---|---|---|---|---|---|---|---|---|
| | *Slow* | | | *Fast* | *Slow* | | | *Fast* |
| *Items per ear per sec.* | 1 *per* 2 *sec.* | 1 *per* 1½ *sec.* | 1 *per* 1 *sec.* | 1 *per* ½ *sec.* | 1 *per* 2 *sec.* | 1 *per* 1½ *sec.* | 1 *per* 1 *sec.* | 1 *per* ½ *sec.* |
| Dichotic lists Naval Ratings .. .. | 68·2 | 60·7 | 64·3 | 53·2 | 74·3 | 75·7 | 78·5 | 89·9 |
| Dichotic lists Lab. subs. .. .. | 97·2 | 90·3 | 81·5 | 75·9 | 92·1 | 88·0 | 83·8 | 88·0 |
| Monaural mixed class lists .. .. | 95·5 | 98·9 | 97·8 | 92·2 | 96·1 | 96·1 | 95·0 | 100 |

If we take the whole-list score first, and compare the performance at the slowest speed (one item/2 sec.) with that at the fastest (one item/½ sec.), we find that of the seven subjects who had the alternating recall condition, five did better at the slow speed and two showed no difference ($p < 0{\cdot}05$ one-tailed.) On the other hand, of the eight subjects who recalled in successive order, six were better at the fast speed and again two showed no difference ($p < 0{\cdot}05$ two-tailed). The difference between the two recall orders in number of subjects doing best at fast and at slow rates is significant by the exact test for a 2 × 2 table ($p = 0{\cdot}01$ two-tailed).

If we take the number of individual digits correct, then in the alternating condition five subjects were better at the slowest speed and one was worse. In the successive condition six were better at the fastest speed and one worse. In a 2 × 2 table the difference is only significant at $p = 0{\cdot}05$ (one-tailed test). If, however, we take the difference between the number of correct items at the slowest speed of presentation and that at the fastest, and compare the two groups of subjects using the Mann-Whitney U test then the difference between the two methods of recall is significant $p = 0{\cdot}004$ (two-tailed).

(*b*) *A.P.R.U. subjects.* It will be recalled that Moray found an *improvement* in the performance with ear-by-ear recall as presentation rate slowed down. Since the Naval Ratings showed the opposite effect, subjects more similar to Moray's were tested in an attempt to resolve the discrepancy; they also spoke responses as his subjects did instead of writing them.

When alternating between the ears these subjects behave in much the same way as the Naval ratings except that their over-all level of performance is higher. Out of six subjects, five do better at the slowest speed than at the fastest and one shows no difference ($p < 0{\cdot}05$ one-tailed by sign test). When recalling ear-by-ear, however, the group shows no difference between the speeds, three doing better at the fastest, two worse and one showing no change. Mean performance is better at the slow apeed, as it was in Moray's paper. Nevertheless, taking the difference between performance (as measured by whole lists or by digits correct) at the slowest speed and that at the fastest, the figure is more positive in the alternating condition in every case ($p = 0{\cdot}032$ two-tailed). Thus for these subjects also, a slow presentation rate confers more of an advantage when alternating between ears than when recalling ear-by-ear.

(2) *Monaural mixed letter and digit lists.* The mean performance for each group is shown in Table I. Subjects having the items presented in alternating order do best at the slowest speed, while subjects having the successive order of presentation do best at the fastest. Comparison of performance at the fastest speed with that at the slowest yields the same figures for both scores; when alternating between the classes four subjects do better at the slow speed and one shows no change. On the other hand, when recalling successively, four subjects do better at the fast speed and one shows the same performance at both speeds.

The difference between the two recall orders in numbers of subjects doing better at fast and at slow rates is significant ($p = 0{\cdot}05$ two-tailed) using the exact test for a 2 × 2 table.

## DISCUSSION

*Dichotic all digit lists.* The results show quite clearly that reducing the rate of presentation does not have the same effect upon both modes of recall and so Broadbent's earlier findings (1954) cannot be explained, as Moray suggested, by a general effect of speed of presentation upon recall.

It is interesting to note the confirmation of Moray's result, that a more intelligent type of subject does inprove his ear-by-ear recall with spoken responses when the items arrive slower. The explanation may be the faster form of response, but there is some reason to think that the effect is to be explained by rehearsal in alternating order during the presentation. Four of our six subjects reported this rehearsal and, as has been commented previously (Broadbent and Gregory, 1961) the relationship of difficulty to serial order in Moray's results supports this interpretation. Whatever the explanation, however, this way of doing the experiment resembles the other in showing that extra time is more useful when the listener has to alternate. Thus Broadbent's original interpretation, that a certain amount of time is needed to shift attention, becomes plausible again.

*Monaural mixed class lists.* Here also alternation becomes relatively easier when more time is available. This strengthens the interpretation that a class of items can be selected just as a sensory channel can, and that extra time is needed to shift this selection. Furthermore Rabbitt (1962) has shown that in the presentation of coloured letters, the colour and the shape of the letter may act as classes for selection. At first sight, however, it is difficult to see how a class of words can be selected, since they can hardly possess a common physical characteristic as a sensory channel does.

Broadbent (1958) thought that this was a possibility, but Moray (1961) has presented evidence against it. It is relevant, however, to consider certain findings and interpretations of Treisman (1960).

She required subjects to shadow (repeat aloud as it was heard) a passage on one ear and ignore another passage presented simultaneously to the other ear. If the two passages were switched over between the ears, the subjects sometimes repeated one or two words from the wrong ear at the break. They would then continue with the new passage on the original ear. The higher the transition probabilities in the original message the more likely the listeners were to show these intruding words.

When discussing her findings Treisman suggests that within the decoding system, where meaning is analysed by matching the incoming signals with some form of dictionary store of known words, some units are permanently more readily activated than others. One such unit might be that corresponding to one's own name. This would explain the finding of Moray (1959). On the other hand, units may have their thresholds for activation temporarily lowered by incoming signals on the basis of conditional probabilities. Within the dictionary, such transition probabilities between units would have been built up by continued use of a language. Thus following the words "seated at a mahogany" the threshold of the unit corresponding to the word "table" would be low. Treisman therefore puts forward a modified filter theory, which supposes that the selective mechanism acts on all words arriving on the ignored channel by attenuating rather than by blocking them; that is, the filter renders these signals less likely to activate dictionary units. If, however, certain of these words on the ignored channel correspond to dictionary units whose thresholds at that time are lowered by transition probabilities, then these words may still be heard. Treisman further suggests that the threshold of the dictionary units might be altered in a similar way by the selection of a verbal category or class of material. All words of a certain class would have their thresholds raised or lowered relative to those of other classes.

Treisman's model thus provides a way in which a class of items might be selected, even although the mechanism is rather different from that which selects a sensory channel. One might speak of this type of selection as a set to make certain responses, rather than a set to react to stimuli from a certain source: and it is in this sense that these results refer to a "response set." The processes involved are, however, central ones, since the overt response does not occur until well after presentation and could hardly be responsible for the effects of presentation rate. Ultimately a more accurate term than "response set" should probably be found for these central processes: but for the moment it may serve to distinguish them from the form of set which concerns, not the nature of the item, but the channel by which it arrives.

Thanks are due to the Royal Navy for the use of subjects.

## References

BROADBENT, D. E. (1954). The role of auditory localisation and attention in memory span. *J. exp. Psychol.*, **47**, 191–6.

BROADBENT, D. E. (1958). *Perception and Communication.* London.

BROADBENT, D. E., and GREGORY, M. (1961). On the recall of stimuli presented alternately to two sense-organs. *Quart. J. exp. Psychol.*, **13**, 103–9.

GRAY, J. A., and WEDDERBURN, A. A. I. (1960). Grouping strategies with simultaneous stimuli. *Quart. J. exp. Psychol.*, **12**, 180–4.

MILLER, G. A. (1956). The magical number seven plus or minus two: some limits in our capacity for processing information. *Psychol. Rev.*, **63**, 81–97.

MORAY, N. (1959). Attention in dichotic listening: affective cues and the influence of instructions. *Quart. J. exp. Psychol.*, **11**, 56–60.

MORAY, N. (1960). Broadbent's filter theory: postulate H and the problem of switching time. *Quart. J. exp. Psychol.*, **12**, 214–20.
MORAY, N. (1961). Perceptual defence and filter theory. *Nature*, **191**, 940.
RABBITT, P. M. (1962). Short-term retention of more than one aspect of a series of stimuli. *Nature*, **195**, 102.
TREISMAN, A. M. (1960). Contextual cues in selective listening. *Quart. J. exp. Psychol.*, **12**, 242–8.

*Manuscript received 10th July, 1963.*

# A RE-EXAMINATION OF THE INTERRELATIONS AMONG MEASURES OF RETENTION

BY

HARRY P. BAHRICK and PHYLLIS O. BAHRICK

*From the Department of Psychology, Ohio Wesleyan University*

The interrelations among measures of anticipation, recognition and savings are examined. It is shown that (*a*) the difficulty level of a recognition task can be above or below that of anticipation for the same material, and (*b*) the slope of retention curves based upon recognition measures may be more or less steep than the slope of curves based upon recall measures of the same material. Previous contrary conclusions reflect the exclusive use of easy recognition tests, and experimental designs in which the degree of learning is much greater for the recognition than for the recall task.

## INTRODUCTION

Systematic study of the interrelations among indicants of performance is a fundamental task for psychology. Without adequate knowledge of these relations the choice of indicants for a particular investigation often remains arbitrary and findings are likely to be overgeneralized or even misinterpreted (Bahrick, Fitts and Briggs, 1957). In their review of studies dealing with the interrelations among measures of retention, Postman and Rau (1957) conclude: "The one fact for which there is substantial experimental evidence is that tests of recognition yield higher scores than do tests of recall." From their own findings they further concluded that "for all practical purposes the method of recognition shows no forgetting at all during the time interval (2 days) used in the present study."

The present study attempts to demonstrate that (*a*) recognition tests may yield scores which are *higher or lower* than recall tests over the same material. (*b*) The slope of retention curves based upon recognition measures may be *more or less* steep than the slope of curves based upon recall measures of the same material. Previous contrary conclusions (Postman and Rau, 1957; Luh, 1922; Miler, 1957) reflect the exclusive use of "easy" recognition tests in which the correct and incorrect alternatives are highly differentiated.

In all earlier investigations the degree of learning *in relation to the threshold* was always much greater for the recognition task than for the recall task, and the slopes of the retention curves reflect this confounding rather than intrinsic differences between the two types of indicants. In another paper (Bahrick, 1964) it has been shown that dichotomous indicants do not directly reflect the course of forgetting, i.e. they do not directly reveal the rate at which associations weaken in time. Rather, they reflect the number of associations passing below a given threshold level per unit of time. This, in turn, depends not only upon how much the associations weakened during the period, but also depends upon how strong the average associative strength was in relation to the threshold at the beginning of the period, and what the variance of associative strength is.

## METHOD

The experiment was divided into two parts corresponding to the two major types of design commonly employed in studies of retention. In Part I an easy recognition group, a difficult recognition group and an anticipation group were trained until they reached a comparable response criterion and their retention was later tested. In Part II an equal number of training trials were administered to all three groups. The advantages and limitations of these two types of design in relation to retention problems have been discussed in another paper (Bahrick, 1964). Part I was further divided into sections A and B; these two sections represented replications of procedure, but with different stimulus material.

### *Part I—Section A*

*Subjects and groups.* Ninety male and female subjects were randomly divided into three groups: (*a*) an anticipation group (A); (*b*) an easy recognition group (ER); and (*c*) a difficult recognition group (DR). The 30 subjects of each group were further randomly assigned to three subgroups of 10 subjects each.

*Materials.* The learning material consisted of seven paired associates. The stimulus member of each pair was an 8 × 8 matrix with one of the 64 cells filled in black, the remainder left blank. The position of the single black cell was chosen randomly from cells *not* comprising the perimeter of the matrix (in order to minimize anchoring effects). The response member for each stimulus matrix was a randomly chosen, three-digit number.

A relatively easy and a relatively difficult recognition test were used. Both consisted of a chart with seven rows and ten columns. Each row corresponded to one of the paired associates. A sample row of the difficult recognition test for Section A is shown in Figure 1. Within each row, the correct stimulus matrix was randomly positioned among four incorrect matrices. The correct paired-associate response was randomly positioned among the four incorrect ones. The easy recognition chart used wrong response numbers which shared no digits with the correct response number, and incorrect stimulus positions which were three cells removed from the correct position. The difficult recognition test contained incorrect response members all of which shared two of the three digits with the correct number, and incorrect stimulus positions only one cell removed from the correct position. None of the incorrect stimulus and response alternatives for a particular association was identical to the correct ones for another association. Within the above constraints the incorrect stimulus and response members were chosen randomly. The charts for both difficult and easy tests were prepared with three different sequences of the seven rows representing the seven paired associates. An automatic slide projector was used to present the paired associates on a screen located 10 ft. at eye level in front of the subject. The stimuli subtended a visual angle of 3° 10′. An exposure device located at table height on an inclined plane directly in front of the subject was used to present the recognition tests. The device exposed each row of the test blank individually by means of seven shutters.

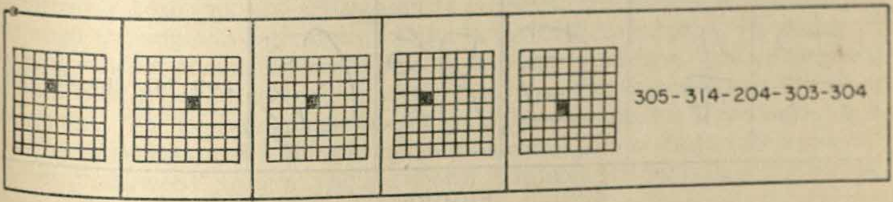


FIGURE 1

One row from the difficult recognition test for matrices.

*Procedure.* Two, three or four subjects serving in the same group were trained simultaneously. They were seated next to each other facing the screen and tape-recorded instructions were presented. They were instructed to observe the material to be displayed on the screen and after the first trial to anticipate each response number by writing it with pencil into the exposed row of an answer sheet located in front of them. The material was presented automatically by showing first the stimulus slide alone for 4 sec. and then presenting the stimulus together with the response number for 4 sec. The exposure device allowed the subject to write his response only during the 4 sec. before the response appeared on the screen. At the expiration of that interval, the shutter closed over the appropriate position on the answer sheet, and the next position was exposed 4 sec. later when a new stimulus slide appeared on the screen. The subjects in the two recognition groups were administered their respective recognition tests during each intertrial interval of 130 sec. The tests were presented on the exposure device one row at a time, and subjects were instructed to make a written indication of correct stimulus and response members on the recognition test sheet. The exposure device presented each row of the test blank for 8 sec., so that a total of 56 sec. was allowed for the test. The balance of the 130-sec. intertrial interval was used as rest period, during which the experimenter scored the recognition tests. The sequence of seven paired associates presented on the screen was randomized on successive trials. Recognition tests presenting the sequence in three different orders were systematically alternated during successive intertrial intervals.

Training was discontinued for subjects of group A following the first trial on which they made seven correct anticipation responses. The subjects in the recognition groups were dismissed following the trial on which they first gave seven correct recognition responses. They were instructed to report for "further training" after 2 hr., 2 days, or 2 weeks, respectively, in accordance with their subgroup assignment.

When subjects reported for their second session, they were given the anticipation or recognition test appropriate to their group under the original conditions of learning. Their training was then resumed under the original conditions until they again reached their original criterion.

*Part I—Section B*

Section B of the experiment replicated section A in every detail of procedure, except that modified Japanese characters were used as stimulus material. A sample row of the difficult and of the easy recognition tests for section B is shown in Figure 2.

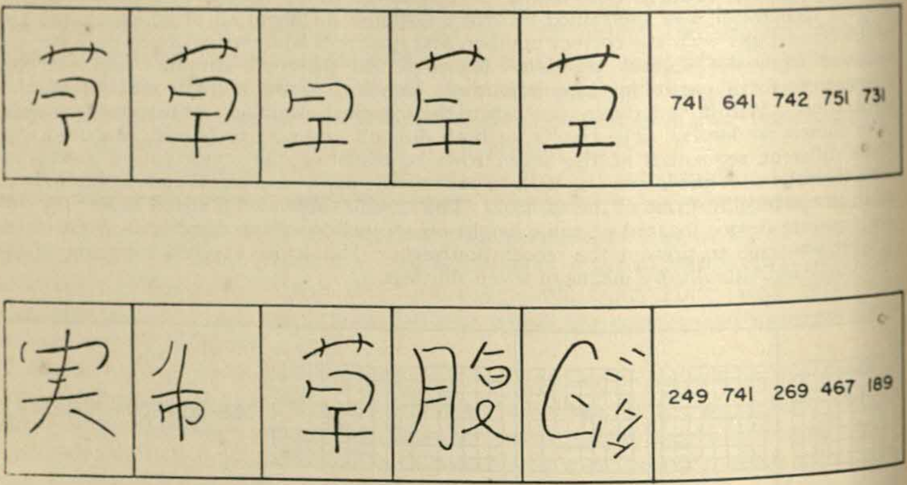


FIGURE 2

One row from the difficult and easy recognition test for characters.

## RESULTS AND DISCUSSION

*Training.* Table I shows the mean number of trials each group required to reach their criterion. Separate analyses of variance were performed for the groups training on matrices and those training on modified characters. Both analyses yielded a significant $F$ ratio ($p < 0{\cdot}01$) between groups, but no significant $F$ ratio ($p > 0{\cdot}05$) between time intervals or for the interaction component of groups × time intervals. The difficulty level of the anticipation threshold fell between that of the easy and difficult recognition thresholds for both types of stimulus material. Thus, the data suggest that recognition thresholds may lie above or below anticipation thresholds depending upon the choice of incorrect alternatives on the recognition test.

### TABLE I

### Learning and Retention Data for Part I

| Condition | *Trials to learn* | | *Percentage retained* | | | | | | *Percentage saved* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | *2 hours* | | *2 days* | | *2 weeks* | | *2 hours* | | *2 days* | | *2 weeks* | |
| | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD |
| M-ER .. .. | 8·0 | 2·8 | 83·7 | 14·9 | 79·2 | 21·4 | 56·9 | 25·9 | 75·3 | 11·9 | 68·4 | 11·9 | 46·4 | 21·7 |
| M-A .. .. | 9·3 | 3·5 | 80·0 | 15·9 | 72·8 | 16·2 | 38·6 | 24·8 | 73·9 | 13·6 | 57·8 | 31·9 | 55·4 | 18·9 |
| M-DR .. .. | 11·5 | 3·2 | 74·7 | 15·7 | 68·8 | 13·5 | 31·7 | 21·4 | 71·6 | 9·2 | 73·5 | 8·8 | 69·6 | 7·4 |
| CH-ER .. .. | 6·9 | 2·5 | 82·2 | 17·8 | 81·0 | 15·7 | 64·3 | 11·4 | 65·9 | 21·1 | 67·5 | 11·6 | 47·9 | 19·3 |
| CH-A .. .. | 14·2 | 4·9 | 71·4 | 16·9 | 64·3 | 17·2 | 21·4 | 16·0 | 71·9 | 12·0 | 66·2 | 30·6 | 43·1 | 21·9 |
| CH-DR .. .. | 15·4 | 4·8 | 67·4 | 14·0 | 56·9 | 19·6 | 36·1 | 23·1 | 70·8 | 11·1 | 63·6 | 19·3 | 65·3 | 13·9 |

*Retention data.*—Table I also shows the mean percentage retained by the recognition and anticipation groups for matrices and characters, respectively. Analyses of variance performed separately for each type of stimulus material show significant differences between groups ($p < 0·01$) for both kinds of material, significant differences between time intervals ($p < 0·01$) for both kinds of material, and no significant interaction component of groups × time interval ($p > 0·05$) for either type of material.

Inspection of Table I shows that the general rule that recognition scores are higher than recall scores does not hold. For both kinds of stimulus material the difficult recognition performance is below performance based on anticipation (one reversal). It is apparent that when training is interrupted at the threshold of each indicant, the amount of forgetting is a function of task difficulty, as determined by the number of trials needed to reach the threshold and independent of the method of measurement. Retention is highest for the ER groups which required the least number of training trials, and lowest for the DR groups which required the greatest number. The anticipation condition, intermediate in difficulty level, yields data which, with the exception of one point, fall in between the ER and DR conditions. This greater amount of forgetting for the more difficult task has been reported previously (Underwood, 1949, pp. 529–31) and can be related to the variable of overlearning. Although training was interrupted at the threshold for all groups, the degree of learning was not completely controlled. Individual associations were overlearned, i.e. practiced beyond the trial on which they were first correctly anticipated or recognized. The average amount of overlearning per association is not equal when training on an easy and a difficult task is interrupted at a common criterion level. Rather, Underwood (1949, p. 529) has shown that more overlearning occurs for the easier of the two tasks.

Data based upon savings scores are also shown in Table I. For the groups training to recognition criteria these data are based upon relearning to a recognition threshold, and for the other two groups, upon relearning to the anticipation threshold.

Analyses of variance performed separately for the two kinds of stimulus material show significant differences ($p < 0·05$) only for the component due to time interval, but not for the variance between groups or for the interaction component of groups × time interval ($p > 0·05$). Thus, the savings scores based upon recognition performance appear comparable to those based upon anticipation performance for the same material, indicating that the savings scores are less sensitive to the previously discussed differential amounts of overlearning associated with the three groups. These overlearning effects produced significant differences among the A, ER, and DR retention curves, but not among the savings scores of these same subjects. Where large amounts of overlearning are involved, we would expect dichotomous indicants and the savings scores based upon them to become equally insensitive. If all associative strength remain above the threshold level after the retention interval, the savings score is 100 per cent. and is thus insensitive to any weakening of the association that may have occurred. When the average amount of overlearning is small, so that many of the associations fall below threshold strength on the retention test, the dichotomous scores are maximally sensitive. The savings scores remain sensitive to continued weakening of the associative strength longer than dichotomous scores, and thus they must allocate less of their total variance from 0–100 per cent. to the time range over which the dichotomous scores are most sensitive. Greater *relative* sensitivity over later time intervals must be achieved at the expense of less *relative* sensitivity over early time ranges, if the total drop of the curve is equal. This accounts generally for the more gentle slope of curves based upon savings scores, in comparison with curves based on anticipation performance. It has been shown, however, that curves based upon savings scores may drop more sharply immediately after learning than curves based upon recognition or free recall indicants, and may later cross these curves (Luh, 1922). Indeed, this is true for four of the six curves in this experiment when they are compared with the corresponding curves based upon the dichotomous indicants (see Table I). Two factors determining relative sensitivity of scores are responsible for this: overlearning with respect to the threshold of the dichotomous indicant, and the number of trials on which the savings score is based. It has already been shown that the largest percentage of overlearning per association occurred under the conditions requiring the fewest trials to learn, and that, as a result, performance based upon the dichotomous indicants showed less initial decline for these conditions. This is not true for performance based upon savings scores. Although a great deal of overlearning would also reduce the sensitivity of this indicant, as has been previously discussed, the opposite effect results from having fewer trials, *per se*, in original training. Trials are the units on which the savings scores are based and when there are not many of them the unit becomes coarse, and a larger initial drop in the curve must be expected. Thus, if an association has just barely dropped below the threshold of the dichotomous indicant, it may take only one trial during relearning to bring it back above threshold. This would constitute an 80 per cent. savings score if the number of trials to learn and relearn are five and one, respectively, but a 90 per cent. savings score if the trials are 10 and 1, respectively. The crossing of curves based upon savings scores with curves based upon dichotomous indicants is, therefore, most likely when there are fewer trials to learn, and this is born out by examining the data in Table I. For the two conditions requiring the fewest trials to learn (M–ER and CH–ER) performance based upon savings scores is below the corresponding performance based on dichotomous indicants and after two weeks no reversal has yet occurred. For the M–A and M–DR conditions requiring more trials during original learning the savings scores are initially below the corresponding dichotomous scores, but cross these during the 2-week period. For the conditions requiring the largest number of trials during learning (CH–A and CH–DR) performance based upon savings scores is above the corresponding dichotomous performance.

None of the data presented so far reveal to what extent the results for the ER and DR groups during training and retention are influenced by successive exposure to recognition tests. These repeated tests may facilitate training and retention, since they involve additional presentations of the correct stimuli and responses.

Even though the tests were administered without knowledge of results they may, nevertheless, strengthen associations, particularly those already above the recognition threshold. At the same time the successive exposure to so many wrong stimuli and responses on the recognition tests may create interference effects which could retard training and lower subsequent retention scores. It is not possible to determine the extent to which interactions among these facilitating and interfering factors affect the data and, therefore, the conclusions arrived at so far cannot be generalized with confidence to situations in which the method of successive recognition tests is not used.

Training to a given criterion of recognition performance necessarily involves the administration of successive tests, and therefore this type of design cannot be used if such tests are to be avoided. Instead it is necessary to use the second major type of design, i.e. to train the various groups for a common number of trials. The principal disadvantage of the latter design in retention studies is that it offers even less control over the degree of original training.

*Part II*

*Subjects and groups.* One hundred and twenty male and female volunteers from undergraduate psychology courses at Ohio Wesleyan University were arbitrarily assigned to an ER group, an A group, or a DR group. The 40 subjects of each group were further arbitrarily assigned to four subgroups of ten subjects each.

*Procedure.* The matrix material previously described was used, and all 120 subjects received eight trials of anticipation training under the same conditions used for the A group during Part I of the experiment. Immediately at the end of training three of the subgroups were given a single ER, DR or anticipation trial, respectively, and were dismissed. The single recognition trial was administered with the same test and under the same conditions previously described. The remaining subjects were instructed to report for "further training" after 2 hr., 2 days, or 2 weeks, in accordance with their subgroup assignment. When they returned, they were also given a single ER, DR or anticipation retention test under the conditions previously specified.

## Results and Discussion

Table II shows the performance of each group on the eighth trial of anticipation training. Analysis of variance reveals no significant differences ($p > 0{\cdot}05$) among any of the 12 sub-groups.

Retention performance is also shown in Table II. Analysis of variance shows significant differences ($p < 0{\cdot}01$) between groups, as well as between time intervals, but no significant $F$ ratio ($p > 0{\cdot}05$) for the interaction component of groups × time intervals.

Both of the major conclusions of Part I are supported by the retention data shown in Table II. After a constant number of training trials, performance on the difficult recognition task is below that on the anticipation task. This clearly suggests that the threshold related to the DR task is above the anticipation threshold. The comparison here is independent of the administration of successive recognition tests, and also controls the last trial learning effect in anticipation training (Underwood, 1954). The retention data clearly show that the slope of recognition curves is not necessarily less steep than the slope of recall curves. Rather it is apparent that the degree of learning in relation to the threshold of the indicant is an important variable. Previous contrary generalizations are based upon the exclusive use of easy recognition tasks, and thus reflect *unilateral* confounding of the degree of training variable with the method of measurement. The present study has not established that recognition and recall measures will yield identical shapes of retention curves when the degree of training and the difficulty level of the tasks are controlled. Such proof of the null hypothesis would be hard to come by, even if one succeeded in holding both of

TABLE II

LEARNING AND RETENTION FOR PART II

|  | 1 *min.* |  | 2 *hours* |  | 2 *days* |  | 2 *weeks* |  |
|---|---|---|---|---|---|---|---|---|
| *Condition* | M | SD | M | SD | M | SD | M | SD |
| *No. of correct responses on the eighth trial of anticipation* |  |  |  |  |  |  |  |  |
| M–ER .. | 5·1 | 1·70 | 5·9 | 1·30 | 5·8 | 1·54 | 4·9 | 1·92 |
| M–A .. | 5·0 | 2·45 | 6·0 | 1·00 | 5·9 | 1·64 | 4·8 | 1·66 |
| M–DR .. | 5·3 | 2·72 | 5·2 | 1·78 | 5·7 | 1·55 | 4·5 | 1·80 |
| *Percentage of correct responses on the retention test* |  |  |  |  |  |  |  |  |
| M–ER .. | 80·0 | 24·1 | 85·7 | 15·7 | 77·1 | 18·3 | 38·6 | 21·2 |
| M–A .. | 78·6 | 28·0 | 62·9 | 25·7 | 65·7 | 28·7 | 35·7 | 29·4 |
| M–DR .. | 54·3 | 25·4 | 28·6 | 15·7 | 41·4 | 18·6 | 14·3 | 20·2 |

the other variables completely constant. By varying the difficulty level of recognition tasks, *both above and below* that of the recall task, the study has shown, however, that previous generalizations regarding differences between the indicants have been misleading.

A reminiscence effect after 2 days is suggested by the data in Table II. This was also observed by Postman and Rau, but was not apparent in Part I of the present study. The effect appears in Part II for the A group, as well as the DR group and, therefore, seems unrelated to the variables of successive recognition testing. It may reflect an interaction of task difficulty and the degree of learning.

The other obvious difference between the data of Tables I and II relates to the decline of retention performance during the 2 hr. after learning. This loss is more pronounced for the groups trained to a threshold criterion, particularly for the ER and A groups. This reflects the inherent characteristics of the threshold-training design in which a much larger portion of associations will be barely above threshold at the end of training and will, therefore, be more likely to drop below threshold during the immediate post-learning interval.

This research was supported by the United States Public Health Service Grant M-5685 to Ohio Wesleyan University. The writers wish to thank Dr. Louis Herman for his many helpful comments, and the following students for their help in collecting and analysis of data: Evan Ehmann, Jean McGovran, Nancy Reynolds, James Saunders, Frank Schneider, Stewart Wheeler, and Richard Weist.

## REFERENCES

BAHRICK, H. P. (1964). Retention curves—facts or artifacts? *Psychol. Bull.*, **61**, 188–94.

BAHRICK, H. P., FITTS, P. M., and BRIGGS, G. E. (1957). Learning curves—facts or artifacts? *Psychol. Bull.*, **54**, 256–68.

LUH, C. W. (1922). The conditions of retention. *Psychol. Monogr.*, **31**, No. 142.

MILER, A. (1957). Vergleich der Vergessenskurven für Reproduzieren und Wiedererkennen von sinnlosem Material. *Zeit. exp. ang. Psychol.*, **7**, 29–38.

POSTMAN, L., and RAU, L. (1957). Retention as a function of the method of measurement. *Univ. Calif. Publ. Psychol.*, **8**, 217–70.

UNDERWOOD, B. J. (1949). *Experimental Psychology*. New York.

UNDERWOOD, B. J. (1954). Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, **51**, 276–87.

*Manuscript received 6th August, 1963.*

# ON THE RELATIONSHIP BETWEEN CODING AND THE ABILITY TO SEE WHAT ONE IS DOING IN A SENSORI-MOTOR TASK*

BY

J. A. LEONARD

*From the Medical Research Council Applied Psychology Research Unit, Cambridge*

In order to examine the possible interaction between different display-control relationships and the ability to look at what one is doing an experiment was carried out under four conditions: half the subjects were trained with a colour code display, the other half with a positional code; within each of these groups half the subjects were allowed to look at their hand while performing, and the others were not. It was found that the effect of not being allowed to look was larger with the colour code in the initial phase of training, and larger with the positional code in the final phase of training.

Interest in three areas led to the experiment to be described: looking, coding and training.

Szafran (in Welford, 1950, pp. 55 *et seq.*) had observed that older subjects tended to persist in looking at what they were doing even when there was no apparent need to do so. While carrying out an experiment with both older and younger subjects on a sensori-motor task described in this *Journal* (Leonard, 1953) it became possible to compare the "looking" pattern of two age groups (Leonard, 1952, pp. 49–61, and 120–2). Older subjects not only appeared to look more frequently at what they were doing than younger ones, but the whole pattern and the possible purpose of this activity appeared to be different. A skilled young subject appeared to give the occasional, sampling kind of glance at his performing hand while keeping the display in view most of the time. Older subjects tended to alternate between looking at the display and looking at their hands, and not to move until they had an opportunity to look. A few years later, when using the same task, but with a more difficult display-control relationship than previously, the writer noticed that young subjects now seemed to behave rather like older ones had done with the simpler display. It appeared that, with the more difficult display, subjects "looked" because they used the control as an aid in translating the displayed information into action.

It seemed clear from the work of Fitts (Fitts and Seeger, 1953) and of others that there might be an important difference between what Crossman (1956) has called symbolic and non-symbolic displays: those where the association between display state and required action has to be learned and those in which the association is "immediately" obvious. One would like to know just what it is that makes the difference and for how long such differences are maintained during controlled periods of practice. It is clear that many apparently simple display-control relationships are simple only for people who have grown up in a given culture. It is also clear that some differences which have been found appear to persist well beyond the immediate learning phase of the conventional choice-reaction time experiment (Leonard, 1961).

Finally there is a case for setting up miniature training situations within the constraints imposed by a laboratory in order to investigate variables which are of both theoretical and practical interest. In the particular instance, apart from the theoretical points mentioned earlier there is some practical interest in the problem of whether keyboard operators should or should not be allowed to look at their hands early in practice: the traditional viewpoint is that to allow looking in the early stages fosters the establishing of bad habits, since final performance requires operation by touch. More recently it has been suggested that no harm is done in allowing students to look at their hands early in training since this avoids a lot of frustration, and most students shed the habit of their own accord (West, 1957).

---

* The substance of this paper was reported at the annual meeting of the British Psychological Society in 1962.

The combination of these interests led to the experiments to be described in which half of the subjects performed with a display in which direction of hand-movement to be made was indicated by an arbitrary assignment of one of five colours and the other half with a display in which direction was indicated by a positional display. Half the subjects in each group performed throughout with a cover over their hand, while the other half could look at their hands all the time. Two questions in particular were to be answered: (1) will the expected difference between the colour and the positional displays disappear with practice? and (2) will there be any interaction between withholding vision and the type of display used?

It seemed possible that the extensive amount of practice to be given might reduce the difference expected between displays to zero, for earlier studies had not gone to the same length in that matter. Concerning the interaction between "looking" and "coding" it seemed reasonable to suppose that the effect of withholding vision would be much more marked on the performance of the subjects performing with the colour code than those performing with the positional code: in one case the display was a direct spatial representation of the control panel over which the stylus had to be moved, hence one could expect direction to be "given" to the subject by the display and the withholding of vision to be a relatively unimportant matter. In the case of the colour the display did not indicate direction of required movement in any direct manner.

## METHOD

*Apparatus.* This was a self-paced serial-choice reaction apparatus with a fixed programme of 50 stimuli wired on a uniselector. This programme was completely balanced with respect to the frequency and digram structure of five stimuli, with the further restriction that no stimulus was repeated without at least one of the other four appearing once in between. The programme could be permuted by means of five single-pole, five-way switches. The uniselector was advanced by one step only when the subject made a correct response. Errors were counted on a post-office type counter, and time per trial of 50 correct responses was recorded on a Standard Electric Timer reading to the nearest 0·01 sec. The programming and scoring side of the apparatus was placed outside a noise-attenuating cubicle.

The display was either a pentagon of five small filament lamps 1 in. apart (the "positional" display) or a ground-glass screen 1 in. in diameter set in the centre of the pentagon, in a recess of about 1⅜ in. in depth (the "colour" display). This screen could be illuminated from the rear by one of five coloured ordinary sized lamps which were housed in a light–box: the illumination provided by each lamp was reasonably uniformly spread over the surface of the screen but no attempt was made to match the colours for brightness. The displays were at about eye level and approximately 18 in. from the centre of the control panel.

The control consisted of a flat panel made of hardwood with brass-discs 1¼ in. diameter let in flush. Five discs formed a pentagon corresponding to the display (but with 3½ in. spacing between centres of discs) and one disc was located in the centre. The subject held in his hand a metal-tipped stylus with wire attached.

When the positional display was used there was a one-to-one matching between display state and control, e.g. the light appearing at 12 o'clock meant that the disc at 12 o'clock was the correct disc. The display-control relationship in the colour-code conditions is shown in Table I. (The clock terminology is only approximate.) When the colour display was used, a mask was placed over the positional display.

TABLE I

Display-Control Relationships for Colour Code

| | |
|---|---|
| Orange indicated the disc at 12 o'clock | (0°) |
| Green ,, ,, ,, ,, 2 ,, | (72°) |
| Red ,, ,, ,, ,, 5 ,, | (144°) |
| Yellow ,, ,, ,, ,, 7 ,, | (216°) |
| Blue ,, ,, ,, ,, 10 ,, | (288°) |

A metal frame was built around the control panel to allow a sheet of cardboard measuring 21 in. × 16½ in. to be placed above the control panel. The cardboard was slanted so that the edge nearest the subject was about 4 in. up from the plane of the control panel and protruding about 6 in. in front. The rear edge of the board was approx. 12 in. above the level of the control. When the board was lifted off, a narrow strip of cardboard 21 in. long was left in the position of the proximal edge of the board. The board prevented the subject from seeing the control-panel without interfering with his movements.

The subject sat in a noise-attenuating cubicle. The experimenter could observe the subject through a "one-way" window, and communicate with the subject by raising his voice.

*Task.* The task was that referred to as condition (1) of the 1953 paper: starting from the centre disc the subject had to slide the stylus to the disc indicated by the display, and then slide the stylus back to the centre, when the next move to be made was indicated by the display. The display conditions were, however, slightly different from those described earlier: on touching the indicated disc the display would become blank; if a wrong disc was touched a highpitched buzzer sounded until the centre-disc was touched again (it was necessary to have a buzzer because failure of the display to become blank might also mean that the subject had failed to reach the indicated disc in the "touch" condition).

To complete a trial the subject had to make 50 correct responses.

*Procedure.* The task was administered to four separate groups of five subjects each under one of four conditions, each subject performing under one of these only. Subject allocation was random. One subject was rejected for performance reasons and will be referred to in the Discussion.

TABLE II

Conditions Used in the Experiment. V = Vision, T = Touch

| *Display* | *Visual guidance* | *Group* |
|---|---|---|
| Colour .. | Permitted | CV |
| Colour .. | Prevented | CT |
| Position .. | Permitted | PV |
| Position .. | Prevented | PT |

The subjects attended for 10 daily sessions spread over two 5-day weeks. During each session each subject was given two blocks of five trials of 50 correct responses, each session 500, and the total number of correct responses per subject was 5,000. There was a break of 1½ min. between blocks.

All subjects were instructed to be as quick and as accurate as possible and no acceptable error level was specified.

Subjects were informed of the time taken and the number of errors made per trial immediately after each trial. The subjects being allocated to perform with the colour display were given a small card as a training aid for the first two trials of the experiment.

This card was fixed to the left of the display as the subject was looking at it and gave the colour names together with the corresponding clock-positions. Subjects allocated to perform with their hand screened from their eyes were allowed and encouraged to look under the screen between trials. They were also allowed to look under the screen if they completely failed to reach the indicated disc. This somewhat flexible criterion was in fact easy to apply as will be seen in the discussion.

The subjects were 20 volunteer ratings from the Royal Navy, aged between 18 and 26 years.

## RESULTS

Figure 1 shows the performance curves for the four groups in terms of mean times for 50 correct responses for each block of 250 responses. Figure 2 shows individual mean times for the whole of the second week's performance.

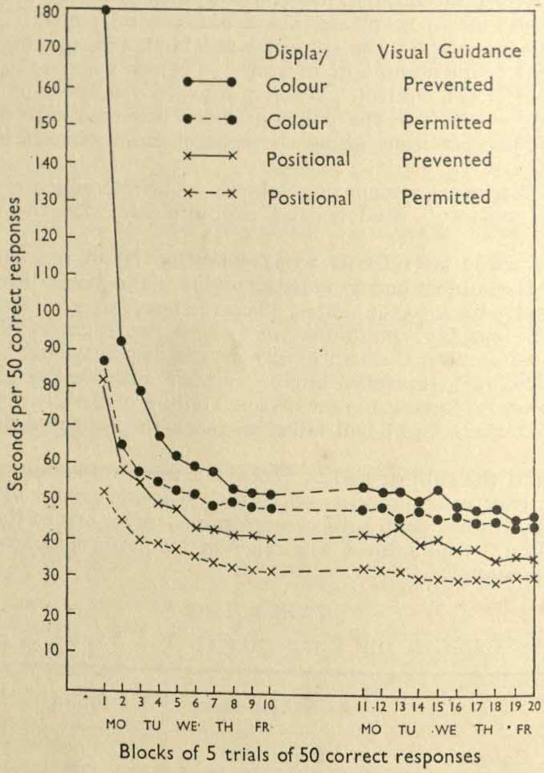


FIGURE 1

Performance curves for the four groups of 5 subjects. Each point is an average of 25 readings.

Errors are not shown because they were very low except for the first few trials on the colour display.

*The effect of displays.* There was a clear difference between the two display conditions, with the two distributions for the second week showing no overlap at all. Table III shows the percentage difference between CV and PV for alternate blocks. There was a fairly consistent percentage difference throughout except for the early phase of training.

*The effect of withholding visual guidance.* While it is clear that on average there is a slowing down of performance due to the withholding of visual guidance, inspection of the performance curves in Figure 1 indicates that there is an interaction with display conditions and stage of practice. There are several ways of describing this interaction:

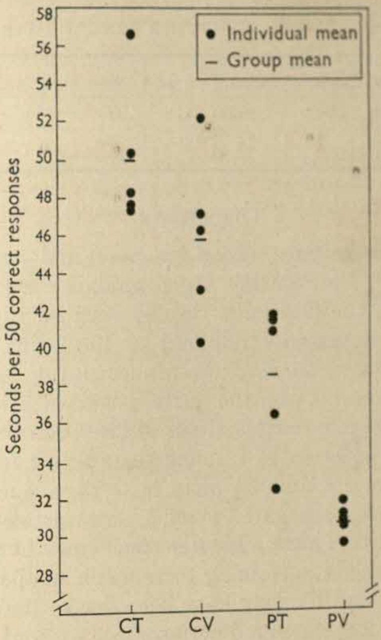


FIGURE 2

Individual Means and Group Means for the whole of the Second Week, in seconds for 50 Correct Responses.

TABLE III

THE DIFFERENCE CV–PV EXPRESSED AS A PERCENTAGE OF PV FOR ALTERNATE BLOCKS

| Blocks | 1–5 | 11–15 | 21–25 | 31–35 | 41–45 | 51–55 | 61–65 | 71–75 | 81–85 | 91–95 |
|---|---|---|---|---|---|---|---|---|---|---|
| | 66·0 | 45·3 | 41·6 | 41·9 | 46·2 | 44·7 | 43·9 | 52·3 | 48·3 | 44·2 |

(1) It would appear that the effect of withholding visual guidance is greater for the colour-display earlier in practice and greater for the positional display late in practice;

(2) The differences between CT and CV, PT and PV, expressed as percentages of CV and PV respectively are shown in Table IV.

(3) Mann-Whitney U-tests were carried out for successive blocks on CT vs. CV and PT vs. PV: from about the eighth block onwards the differences between CT and CV ceases to be significant, while that between PT and PV is never less significant than at the 4 per cent. level. The hypothesis which stated that the effect of withholding visual guidance would be more marked with a

colour display than with a positional one is therefore only verified for the early stages of practice.

TABLE IV

THE DIFFERENCE T–V EXPRESSED AS A PERCENTAGE OF V FOR ALTERNATE BLOCKS
TOP LINE COLOUR DISPLAY, BOTTOM LINE POSITIONAL DISPLAY

| *Blocks* | 1–5 | 11–15 | 21–25 | 31–35 | 41–45 | 51–55 | 61–65 | 71–75 | 81–85 | 91–95 | Display |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 106·8 | 38·3 | 19·1 | 20·7 | 9·41 | 13·5 | 14·1 | 16·1 | 8·1 | 5·1 | Colour |
| | 56·5 | 40·5 | 29·5 | 24·7 | 24·5 | 24·9 | 26·8 | 33·9 | 25·7 | 17·3 | Positional |

## DISCUSSION

*The interaction between displays, visual guidance, and phase of training.* It was expected that the effect of preventing visual guidance would be greater with the colour display than with the positional display because the latter provided direct information about the movements required of the hand. Confirmation of this expectation would have gone some way towards accounting for the difference between the two display conditions. Only the early phase of training provided such a confirmation and there are two complications of interpretation.

Firstly, the fact that early on in training there was a larger difference between CT and CV than between PT and PV may have been due to the subjects in CT having to learn the colour code and "touch" performance simultaneously. Both variables taken by themselves have a greater effect early in training.

Secondly, the fact that late in training there was a smaller difference between CT and CV than between PT and PV may have been due to the fact that the characteristics of the colour display used favoured performance without visual guidance. With the positional display subjects might have been able to interprete a display change out of the corner of their eyes while glancing down at their hands. The CT–PT comparison may be taken as a possible control for the difference between display conditions not being due to this "artefact." This is of some importance because the same criticism could be raised about other comparisons between symbolic and non-symbolic codes.

*The difference between the two display conditions.* The expected difference between the positional and the colour display was maintained throughout. In a further series of experiments a similar difference was found between a positional and a numerical display (Leonard and Newman, 1965). It seems as if very much longer periods of training would be required before the "symbolic" catches up with the "non-symbolic."

*Learning to perform with colour code and without visual guidance.* By contrast, subjects on the whole did not find it difficult to "learn" those two features of the experiment which are more traditional topics of laboratory learning. It was rare for subjects performing with the colour code to make many mistakes after the first day and the large differences in performance due to the absence of visual guidance disappeared after the second day. One is tempted to think of this kind of learning, and the learning which may eventually enable a subject to turn a symbolic code into a non-symbolic one, as different in kind.

During the early phase of performing without visual guidance subjects failing to reach an indicated disc after several attempts would apparently "freeze" up. They were then allowed to look under the cover and generally one or two such corrections were all that was needed for the phenomenon not to happen again. The one subject whose data were not included here continued to experience this difficulty throughout.

The very large differences due to the absence of visual guidance found in the early phase lends some support to those keyboard instructors who allow their students to look at their hands to avoid attrition through boredom during the first weeks of keyboard training (West, 1957).

*The development of a "miniature training situation."* The task reported on here was not found suitable for this purpose mainly because of its lack of face-validity. A keyboard task has been developed since (Leonard and Newman, 1965).

## REFERENCES

CROSSMAN, E. R. F. W. (1956). The information-capacity of the human operator in symbolic and non-symbolic control processes. In DRAPER, J., Ed. *The Application of Information Theory to Human Operator Problems.* W.R.(D) Report No. 2/56 Ministry of Supply.

FITTS, P. M., and SEEGER, C. (1953). S–R compatibility: spatial characteristics of stimulus-response codes. *J. exp. Psychol.*, **46**, 199–210.

LEONARD, J. A. (1952). *Some Experiments on the Temporal Relation between Information and Action.* Ph.D. dissertation, Cambridge.

LEONARD, J. A. (1953). Advance information in sensori-motor skills. *Quart. J. exp. Psychol.*, **5**, 141–9.

LEONARD, J. A. (1961). Choice-reaction time experiments and information theory. In CHERRY, E. C., Ed., *Information Theory*, London, 137–45.

LEONARD, J. A., and NEWMAN, R. C. (1965). On the acquisition and maintenance of high speed and high accuracy in a keyboard training task. *Ergonomics* (in press).

WELFORD, A. T. (1950). *Skill and Age.* London.

WEST, L. J. (1957). Review of research in typewriting learning with recommendations for training. *U.S.A.F. Pers. & Training Res. Rep.*, AFPTRC-TN-57-69.

*Manuscript received 14th October, 1963*

# THE COMBINATION OF INFORMATION FROM DIFFERENT SOURCES

BY

R. DAVIS

*From the Department of Psychology, The University of Reading*

In this experiment subjects were presented with visual information from two different sources and were required to combine it in order to make the correct response. The time interval between the two signals was varied in two different ways (*a*) regularly and (*b*) randomly. Reaction times were measured from the moment of occurrence of the second signal. By this means the time course of the decision procedure involved in combining the information from the two sources was analysed.

Results indicate that subjects may deal with the situation in two ways (i) by means of a "perceptual" classification in which the individual elements are not analysed separately, or (ii) by means of an "intellectual" classification in which each signal is analysed sequentially. These two methods correspond to the experimental conditions of (*a*) regular intervals and (*b*) random intervals.

It is argued that when subjects use the latter strategy the results are consistent with the conception of the human operator as an intermittent analysing system.

## INTRODUCTION

The notion of channel capacity derived from information theory has had fairly wide application in the study of human sensory motor performance. It has been used as a theoretical background, for example, in experiments on discriminative capacity, immediate memory and reaction time. In most, though not all, of these situations the subject's task has been to preserve a correspondence between stimulus array and response array. That is, stimuli and responses have been related in a one-to-one manner and the subject has had to make a straightforward translation from stimulus set to response set. This is true of most of the experiments on absolute judgements (e.g. Miller, 1956), experiments on multichoice reaction times (e.g. Hick, 1952; Hyman, 1953) and experiments on stimulus-response compatibility (e.g. Fitts and Seeger, 1953).

There are, however, situations which have not been quite so well explored, in which the number of response categories is less than the number of stimulus categories, so that the subject does not act merely as a translator or conserver of information, but, in a sense, has to reduce the amount of information presented to him into a smaller number of categories.

This might be regarded as a very elementary description of concept formation as Garner (1962) points out, and Posner (1962) has applied the idea of information reduction to an analysis of simple thinking behaviour.

In these situations the response uncertainty limits the amount of information transmitted, and the question arises as to whether this is the sole determinant of response time or whether variations in stimulus uncertainty play any role. Several experimenters, e.g. Bricker (1955), Rabbitt (1959), Broadbent and Gregory (1962), have produced evidence suggesting that response uncertainty is of overriding importance and have argued that stimulus uncertainty has little effect.

On the other hand Crossman (1953) showed that in a situation where playing cards were sorted into two response categories, black pictures with red numbers into one pile and red pictures with black numbers into another pile, the time taken was more closely related to stimulus uncertainty than to response uncertainty (or information transmitted). Furthermore, Posner (1962) has shown that the quantity most closely related to task difficulty in a variety of forced paced tasks was not the response uncertainty (or information transmitted) but the amount of information reduction from stimulus to response categories.

It is, however, necessary to distinguish between two types of situation in which reduction of information takes place.

(1) In which information irrelevant to the basis of classification is merely excluded, e.g. where stimuli vary in both colour and shape, but the classification is made on the basis of one dimension of variation only.

(2) In which several dimensions of variation in the stimulus material must be taken into account in order to determine the correct response, e.g. a classification of red circles, red squares, black circles, black squares into two categories:

Red circle or black square/black circle or red square—a task which is similar to Crossman's (1953). In this case, both colour and shape must be taken into account, but although all the information in the stimulus situation must be used to determine the correct response, the characteristics of any particular stimulus cannot be recovered merely by *a posteriori* knowledge of the correct response made to it.

The first situation might legitimately be described as "filtering" in the sense of discarding irrelevant information. The second situation is not mere filtering. A rather more complex type of intellectual operation is taking place, as contrasted with the perceptual classification in the first situation.

In practice it may not be possible to distinguish clearly between these two situations. The response times obtained by Morin, Forrin and Archer (1961) suggest that with practice, subjects may show a gradual transition from an "intellectual" to a "perceptual" classification. However, the way in which this might happen is not fully understood and it would seem that any theoretical analysis should start by distinguishing between the two situations.

The present experiment deals with a situation in which four stimulus categories are reduced to two response categories, using all the stimulus information, but instead of having two dimensions of variation, the stimulus information is provided consecutively from two different sources, and the subject's task is to combine the information from the two sources in order to perform the appropriate response. Thus, two sources I and II can each generate one or other of two alternative signals, A or B, giving rise to four possible stimulus configurations, AA, AB, BA, BB. The subject's task is to decide whether the two signals are the same or different and to press an appropriate key. This is similar to the second situation discussed above, where all the stimulus information has to be used, and one would therefore expect the time taken for this very simple intellectual operation to correspond to the time for a choice among four alternatives along a single dimension.

If however, the signal from the second source is not given simultaneously with the first, but at some short time afterwards, one can examine other possibilities. Thus with a sufficiently long time interval between signals, the subject will have already analysed and classified the first signal before the second signal arrives, and the number of possible stimulus configurations will have been reduced from four to two, e.g. if the first signal is A there are only two possible stimulus configurations AA and AB, to decide between when the second signal arrives, so that one would expect the time taken to respond, measured from the moment of occurrence of the second signal, to correspond to the time taken to decide between two alternatives on a single dimension.

Hence, as long as the interval between signals is long enough for the analysis of the first to be completed by the time the second arrives, there ought to be a relatively constant two-choice time between the onset of the second signal and the response. However, if one supposes that the central analysing systems behave in an intermittent manner as suggested by much previous evidence (Hick, 1948; Welford, 1952, 1959; Davis, 1957, 1959, 1962) there will be no advantage in the signal from the second source arriving before the analysis of the first is completed—it will be held in store until the central systems are ready for it. Thus as the interval between signals becomes progressively less than the critical time for the analysis of the first signal, there will be progressively increasing delays in the reaction time, measured from the moment of occurrence of the second signal, and as the interval approaches zero the reaction time will approach the value for a decision between four alternatives along a single dimension.

By varying the interval between signals in this manner one can investigate the time course of this very simple decision procedure.

## METHOD

*Main experiment*

The apparatus was a Dodge-type two field tachistoscope adapted for reaction time measures.

The stimulus card for the first field carried either X or O and similarly the stimulus card for the second field carried either X or O. These were drawn 2·5 cm. square in black ink on white card and were exposed at a distance of approximately 50 cm. immediately in front of the subject, who viewed them binocularly. The first stimulus was always exposed above the second, to avoid possible masking effects between signals, and the vertical separation between the positions of the two signals was approximately 5 cm., measured from centre to centre.

The subject was equipped with two Morse type keys and was instructed to press the left key if the signals were the same and the right key if they were different. He was told to respond as quickly as possible and reaction times from the start of the second signal were recorded to the nearest millisec. using a Venner clock. The subject was not informed of his reaction times, but if he pressed the wrong key (or both keys) a buzzer sounded as an error signal.

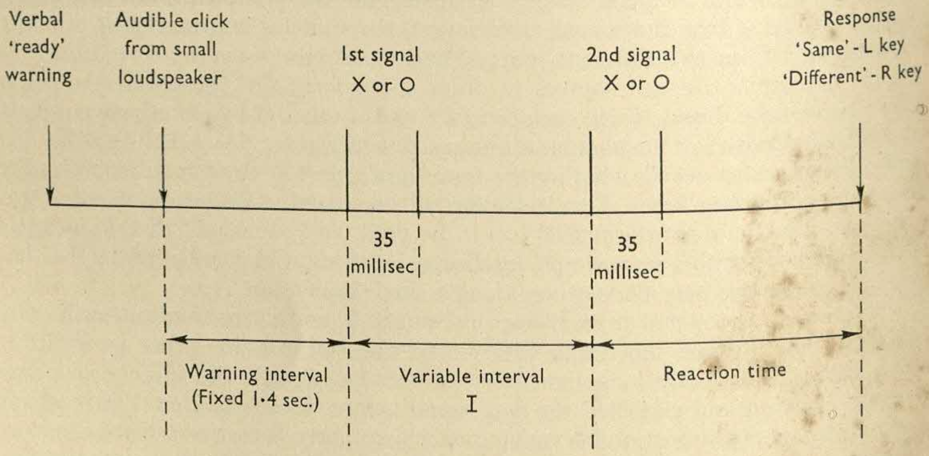

Verbal 'ready' warning
Audible click from small loudspeaker
1st signal X or O
2nd signal X or O
Response 'Same' - L key 'Different' - R key
35 millisec
35 millisec
Warning interval (Fixed 1·4 sec.)
Variable interval I
Reaction time

FIGURE 1
Experimental sequence.

The temporal sequence of events is shown in Figure 1.

The values of the variable interval between signals were 35, 160, 380, 600 and 810 millisec. Thus at the shortest interval of 35 millisec. the onset of the second signal coincided with the switching off of the first signal with no dead time in between. This interval was used as the nearest approach to simultaneity that could be conveniently obtained with the apparatus. The other intervals were also chosen on the basis of convenience in using the available timing system.

The subjects were student volunteers, two men and two women, all right handed, between the ages of 18 and 21, at the University of Paris. They each attended for one preliminary session and six experimental sessions of ¾ to 1 hr. each.

The interval between signals was varied in two different ways. For two subjects (one man, one woman) a *regular interval* condition was used.

These subjects had a block of 12 pairs of signals, all with the same inter-signal interval, and then changed to another value of interval. Two blocks of each interval were presented at each session and the five values of interval were randomized between blocks. Thus, over the six sessions for each subject, each interval was presented 144 times and each pair of signals 36 times at each interval.

For two subjects (one man, one woman) a *random interval* condition was used.

Each successive interval was picked at random from the five possible values, with the restriction that within a block of 20 all intervals occurred with equal frequency. Each subject was presented with five blocks of 20 pairs of signals at each session, so that over the six sessions each interval was presented 120 times, and each pair of signals 30 times at each interval.

In both conditions the actual pairs of signals, XX, XO, OX, OO, were randomized with equal frequency over each block.

If the subject made an incorrect response, the particular stimulus-interval combination was repeated at the end of the block. Mistakes were so few (see Table I) that this procedure had very little effect on the frequency distribution of signals.

TABLE I

MAIN EXPERIMENT

Mean reaction times and standard errors in millisec. for different values of interval I

| | *Subject* | *N* | *I* | *M* | *S.E.* | *No. of mistakes* |
|---|---|---|---|---|---|---|
| Regular Intervals | A | 144 | 35 | 367 | 5·1 | 8 |
| | | | 160 | 381 | 5·5 | 5 |
| | | | 380 | 329 | 4·9 | 2 |
| | | | 600 | 328 | 5·6 | 3 |
| | | | 810 | 335 | 6·9 | 0 |
| | B | 144 | 35 | 379 | 9·5 | 2 |
| | | | 160 | 396 | 7·3 | 3 |
| | | | 380 | 351 | 6·4 | 7 |
| | | | 600 | 366 | 7·5 | 4 |
| | | | 810 | 390 | 8·1 | 1 |
| Random Intervals | C | 120 | 35 | 550 | 9·6 | 5 |
| | | | 160 | 403 | 7·1 | 2 |
| | | | 380 | 363 | 6·1 | 0 |
| | | | 600 | 374 | 6·8 | 0 |
| | | | 810 | 376 | 6·8 | 2 |
| | D | 120 | 35 | 499 | 5·6 | 4 |
| | | | 160 | 430 | 4·1 | 2 |
| | | | 380 | 403 | 4·8 | 2 |
| | | | 600 | 410 | 4·9 | 1 |
| | | | 810 | 439 | 4·3 | 3 |

*Control experiment*

A control experiment was carried out in order to eliminate possible interpretations of the findings, which are discussed below.

In the control experiment another group of four French student volunteers in the same age group were used.

The conditions were the same as those of the main experiment, with the exception that the first stimulus field was blank, and the subjects were called on to make a straightforward two-choice reaction to the second signal, that is, to press the left key for an X and the right key for an O. The interval between the blank first field and the second stimulus field was varied in exactly the same way as in the main experiment. Two subjects performed under a *regular interval* condition, and two subjects under a *random interval* condition.

In this experiment 20 responses for each subject at each interval were taken in a single experimental session.

## RESULTS

### *Main experiment*

A preliminary analysis showed that there were no significant differences between reaction times for "same" or "different" responses, so that results have been pooled to show (*a*) the mean RTs for each subject for 144 responses at each interval in the *regular interval* condition and (*b*) the mean RTs for each subject at each interval in the *random interval* condition. The values of means and standard errors together with the number of mistakes, are shown in Table I and the variations of mean reaction times with interval under the two conditions are illustrated in Figure 2.

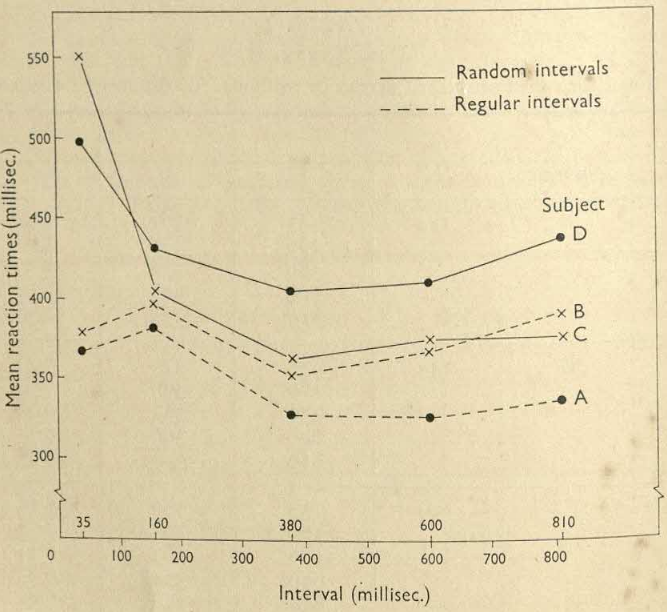


FIGURE 2

*Main experiment.* Variation in mean reaction times with interval, for the regular and random interval conditions.

### *Control experiment*

The mean RTs and standard errors for the 20 responses made at each interval by the four subjects under the two conditions of the control experiment are shown in Table II.

TABLE II

Control Experiment

Mean reaction times and standard errors in millisec. for different values of interval I

| | *Subject* | *I* | *M* | *S.E.* |
|---|---|---|---|---|
| Regular Intervals | E | 35 | 350 | 27·2 |
| | | 160 | 356 | 11·9 |
| | | 380 | 360 | 12·1 |
| | | 600 | 336 | 10·4 |
| | | 810 | 332 | 18·0 |
| | F | 35 | 432 | 21·0 |
| | | 160 | 445 | 14·9 |
| | | 380 | 445 | 11·1 |
| | | 600 | 434 | 11·7 |
| | | 810 | 436 | 12·0 |
| Random Intervals | G | 35 | 363 | 12·2 |
| | | 160 | 336 | 9·9 |
| | | 380 | 352 | 15·9 |
| | | 600 | 368 | 13·6 |
| | | 810 | 393 | 14·6 |
| | H | 35 | 415 | 15·2 |
| | | 160 | 406 | 11·1 |
| | | 380 | 382 | 14·5 |
| | | 600 | 399 | 11·4 |
| | | 810 | 408 | 13·4 |

N = 20 in all cases.

## Discussion

From Figure 2 it is evident that the results from the random interval condition confirm expectations fairly well. For intervals of 380 millisec. and greater, the values of reaction time are similar to the two-choice values found in the control experiment. For intervals less than 380 millisec., reaction time progressively lengthens as the intervals shorten, reaching a maximum (550 millisec. for Subject C, 499 millisec. for Subject D) at the shortest interval of 35 millisec. This is to be expected if the analysis of the second signal does not take place until the analysis of the first signal is accomplished.

The data is somewhat inadequate in two respects:

(1) There is no truly simultaneous presentation of the two signals and this complicates any direct comparison with a four choice control. However, the difference between reaction times for nearly simultaneous presentation (interval 35 millisec.) and reaction times at interval 380 millisec. is of the same order of magnitude as the difference between four-choice and two-choice situations, as shown by other experimenters, e.g. Hick (1952), Hyman (1953).

(2) A finer variation of interval between 0 and 380 millisec. is required in order to define the slope of the reaction time function between these two points.

However, despite these limitations, the general relationship is fairly clear from these results. The results from the *regular interval* condition are similar, except for the very striking departure at the shortest interval of 35 millisec. when the two signals were presented almost simultaneously.

This suggests that, if the two signals appear nearly simultaneously and the subject expects this to happen, then he does not proceed with the two stage analysis, but can respond to the total stimulus pattern by means of a perceptual classification, "same/different" in which the individual elements composing the pattern are not classified separately. The actual signals used would seem to lend themselves quite well to this type of classification although, at present, there appears to be no quantitative way of determining independently the relative ease or difficulty of such a classification. This contrasts with the *random interval* condition, for when the two signals appear nearly simultaneously in this condition the subject cannot know in advance that this will happen. In this case he immediately starts the analysis of the first signal, and once this has commenced he cannot deal with the second signal until the first analysis is completed.

Thus, two methods of dealing with nearly simultaneous signals appear as different strategies imposed by the experimental conditions.

A number of other points deserve discussion.

The first is whether any peripheral masking between signals at the shortest interval may have affected the results. The two signals were arranged to occur in different spatial positions to avoid this possibility and the results from the *regular interval* condition make clear that such masking could not have had any appreciable effect.

Secondly: an alternative explanation for the results might be proposed, similar to an argument which has been used in an attempt to explain delays at short intervals in simple reaction time situations. It may be put in this form. The reason why the subject shows a delay in responding when the inter-stimulus interval is short is not because the central systems are occupied with the analysis of the first signal, but merely that the subject is unprepared for any signal following at such a short interval. The results of the main experiment might be thought compatible with this type of explanation, but it is rendered untenable by a comparison with the control experiment. In the two conditions of the control experiment the arrangement of the inter-stimulus intervals was exactly the same as in the main experiment. The only differences were that the first stimulus field was blank and the subject made a straightforward two-choice response to the signal presented by the second stimulus field. If there were an effect due to the subject's unpreparedness, then it should also have appeared in the control experiment. No such effect appears. The reaction times show very little variation with the interval between blank field and stimulus field over the whole range of intervals used. Hence the lengthening of reaction times in the *random interval* condition of the main experiment must have been due to the presence of an actual signal which had to be classified by the subject.

Thirdly: the author does not wish to argue that there is absolutely no effect due to the temporal uncertainty of the second signal. There is evidence for an optimal inter-stimulus interval about the middle of the range of intervals used, which suggests that as the interval becomes longer the subject is less able to adjust his state of maximum preparedness to the precise moment at which the second signal will occur. However, it should be clear that the magnitude of this effect is relatively small compared with the lengthening of reaction times in the *random interval* condition of the main experiment, and that, for the reasons given, this lengthening could not have been produced merely as a consequence of temporal uncertainty.

The most likely explanation seems to be that the increased reaction times are due to the subject having to analyse and classify the first signal, and the results support the intermittent, or single channel, conception of the human operator, in the sense that if the central systems are occupied with the analysis of any signal, the analysis of any subsequent signal will be delayed until the first is completed.

The experiments were carried out while the author held a visiting fellowship from the Department of Scientific and Industrial Research at the Laboratoire de Psychologie expérimentale de la Sorbonne, and he would like to thank Professor P. Fraisse for providing experimental facilities. The data was analysed and further experiments are proceeding as part of a project on the presentation of information in skill situations, financed by the Department of Scientific and Industrial Research at the University of Reading.

## REFERENCES

BRICKER, P. D. (1955). The identification of redundant stimulus patterns. *J. exp. Psychol.*, **49**, 73–81.

BROADBENT, D. E., and GREGORY, M. (1962). Human responses to classes of stimuli. *Nature*, **193**, 1315–6.

CROSSMAN, E. R. F. W. (1953). Entropy and choice time. *Quart. J. exp. Psychol.*, **5**, 41–51.

DAVIS, R. (1957). The human operator as a single channel information system. *Quart. J. exp. Psychol.*, **9**, 119–29.

DAVIS, R. (1959). The role of "attention" in the psychological refractory period. *Quart. J. exp. Psychol.*, **11**, 211–20.

DAVIS, R. (1962). Choice reaction times and the theory of intermittency in human performance. *Quart. J. exp. Psychol.*, **14**, 157–66.

FITTS, P. M., and SEEGER, C. M. (1953). S–R compatibility: spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, **46**, 199–210.

GARNER, W. R. (1962). *Uncertainty and Structure as Psychological Concepts.* New York and London.

HICK, W. E. (1948). Discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, **1**, 36–57.

HICK, W. E. (1952). On the rate of gain of information. *Quart. J. exp. Psychol.*, **4**, 11–26.

HYMAN, R. (1953). Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, **45**, 188–96.

MILLER, G. A. (1956). The magical number seven, plus or minus two. *Psychol. Rev.*, **63**, 81–97.

MORIN, R. E., FORRIN, B., and ARCHER, W. (1961). Information processing behavior: the role of irrelevant stimulus information. *J. exp. Psychol.*, **61**, 89–96.

POSNER, M. I. (1962). An informational approach to thinking. *Univ. Michigan, Dept. Psychol., Tech. Rep. AFSOR*-2635, 02814-9-*T*.

RABBITT, P. M. A. (1959). Effects of independent variations in stimulus and response probability. *Nature*, **183**, 1212.

WELFORD, A. T. (1952). The "psychological refractory period" and the timing of high speed performance; a review and a theory. *Brit. J. Psychol.*, **43**, 2–19.

WELFORD, A. T. (1959). Evidence of a single channel decision mechanism limiting performance in a serial reaction task. *Quart. J. exp. Psychol.*, **11**, 193–210.

*Manuscript received 25th October, 1963*

# THE EFFECTS OF CONTEXT UPON SPEED OF READING, EYE MOVEMENTS AND EYE-VOICE SPAN

BY

JOHN MORTON

*From the Department of Psychology, Reading University**

Subjects read aloud 200-word passages of statistical approximations up to the 8th order. Their eye-movements were recorded together with a trace of the speech output. Speed of reading, using the syllable as the unit of measurement, increased up to the 5th order for slow readers. Fast readers, on the other hand, further increased their speed to the 6th order. This result had been predicted from a hypothesis that fast readers use contextual cues more efficiently.

Measures of the material in the eye-voice span showed an increase up to the 8th order. Fast readers had a larger material span than slow readers beyond the 5th order, a result paralleling the differences in speed increase. It is suggested that the eye-voice span measured in time is dependent upon the chosen reading speed and the material span.

From the eye-movement records there was no variation of the mean duration of fixation between passages, or between fast and slow readers. The average value was about 240 millisec. The number of both forward and regressive eye movements decreased with increase in contextual constraint up to the 5th or 6th order, i.e. with increase in speed of reading. Fast readers were superior to slow readers in both these respects. The effect of decreasing contextual constraint was to produce more regressive movements together with a slightly smaller mean saccade.

## INTRODUCTION

It is known that the reading efficiency of most literate adults can be improved without training. All that is necessary, in many cases, is that the person reads "more actively" and increases concentration (Morton, 1959, 1961). This implies the existence of some potential relevant to the processes of reading which is normally not fully utilized in the untrained adult. The hypothesis is put forward that this potential is related to our knowledge of the statistical properties of the language. If we utilized this knowledge more completely while reading, then the immediate stimulus would become more predictable. It has been shown (Morton, 1964*a*) that the visual duration threshold for a word in a sentence context is an inverse function of the probability of the word in the context. Thus, increasing the probability of a stimulus would make the stimulus more easy to perceive, and, if this were a limiting factor in reading, would result in an increase in reading speed, with no comprehension loss.

The suggested increase in the ease of perception might be utilized either by spending less time on each fixation or by making fewer fixations per line. In the latter case, words could be recognized in spite of being on the periphery of vision.

An alternative formulation of the hypothesis would be: assuming that most individuals do use contextual cues to a certain extent in reading, then the effect of increasing the amount of contextual constraint in a passage would be to increase the individual's speed of reading. If then we have a series of passages with increasing amounts of constraint, the speed of reading should increase up to the passage which has the amount of constraint normally used by the reader. In addition, if faster readers normally use more context, the effect will have its limit for them at a higher order of statistical approximation to English than for the slow readers.

---

* Now at the Medical Research Council Applied Psychology Research Unit, Cambridge.

Sumby and Pollack (1954) found that speed in typing, and speed, errors and mean number of glances per word in reading aloud were all related to the order of approximation to English of the passages used. The present experiment is in part a replication of this work and in part an extension.

## MATERIALS

The passages used in the experiment were 200-word statistical approximations to English. These were constructed by the method due to Shannon (1951) and Miller (Miller and Selfridge, 1950) and typed on quarto paper in pica. The passages used were the Zero, 1st to 6th and 8th order passages. The first hundred words in the 2nd, 4th, 6th and 8th order lists are due to Taylor and Moray (1960) and were used by their kind permission. The subjects who generated the remaining words were undergraduate and senior members of Reading University.

## APPARATUS

In addition to examining the relation between the order of approximation of a passage and the speed of reading for that passage, changes in eye-movements during the reading were studied. What was required was to discover whether the change that took place in the eye-movements as the speed increased consisted of a change in the number of fixations or regressions, or a change in the duration of fixations.

Eye-movements were recorded by the technique of electro-oculography, described by Shackel and his co-workers (Shackel, Sloan and Warr, 1958; Shackel, 1960) among others. This technique utilizes the presence of a standing potential difference between the front and back of the eye-ball. The field from this dipole moves as the eye-ball rotates, and electrodes placed on the skin surface will detect the resulting change in d.c. potential which is then amplified and recorded on paper as a continuous record of the position of the eyes. The electrodes used were of the suction type (Shackel, 1958).

The electrical system is shown in schematic form in Figure 1. The preamplifier was a balanced chopper d.c. preamplifier, designed by Dr. E. R. F. W. Crossman; the main amplifier was a Southern Instruments No. 357 and the two-channel pen recorder a Southern Instruments Pen Oscillograph. The filter was a simple three-stage cascade filter made by the writer and optimized empirically to cut down interference without having too long a time-constant. Only the horizontal eye-movements were recorded; the other channel of the pen recorder was used to record the impulses from a throat microphone, so that the time relationship between eye and voice might be studied. In addition the subject's voice was recorded on a Stenorette recorder so that all errors could be noted.

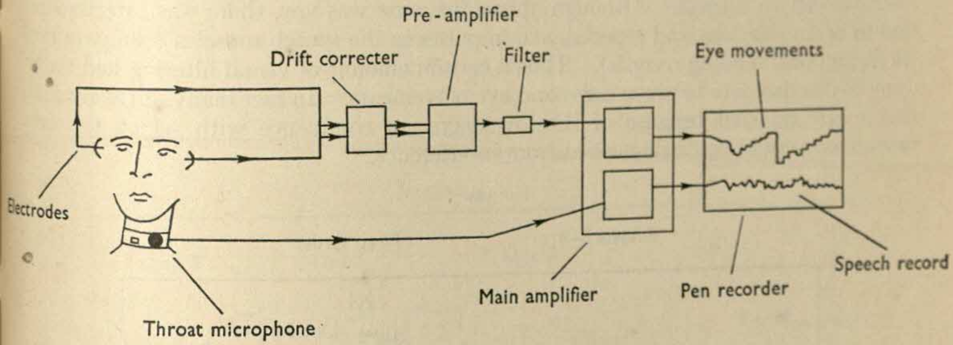


FIGURE 1

Apparatus.

The subject's forehead rested on a padded, rigid bar, with his eyes 35 cm. away from the text, the position of which was adjusted so that it could be read with comfort. The subject's head was not otherwise fixed, since the apparatus in its existing state was prone to drift in any case (this varied with the subject, the electrodes and the experimenter's competence). Any attempt to plot precisely the fixation points would have been at the expense of making the subject more uncomfortable, and the reliability of such assignments would have been low. Head movements were rare and could easily be discriminated from eye-movements on the record.

Thirty-two male subjects were used, senior and student members of the university.

## PROCEDURE

The electrodes and throat microphone were fitted to the subject, who was made as comfortable as possible in the reading position. The subject then fixated a card on the reading stand and the apparatus was adjusted so that fixations at the edge of the card, equivalent to the length of line of the passage, caused the pen to cover about three quarters of the width of the recording paper. To accustom the subject to the apparatus, and vice versa, he was asked to read aloud a 450-word passage from the Pelican *Economics of Everyday Life*. Then the passages of approximation were presented, according to four 8 × 8 Latin Square designs, with instructions to read them "as quickly as possible, minimizing errors." The subject first fixated three calibration points at the top of the paper, then he read the heading before going straight on to the passage.

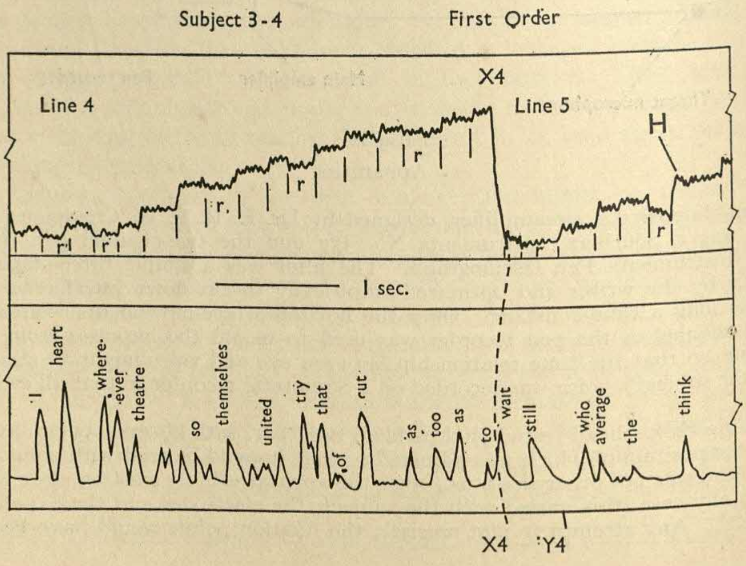
Subjects were told to read the passages without trying to give them any meaning, and were warned that they were all nonsense passages. After the eight lists had been read, a further text from the same book was presented, 444 words long, with instructions to read it, as far as possible, without regard to punctuation or meaning. Many of the subjects found this instruction hard to obey.

The experiment took anything from 20 min. to an hour, depending on the speed of reading, and on whether the apparatus functioned smoothly.

## RESULTS AND DISCUSSION

Sample records are shown in Figure 2. Eye-movements could be distinguished on the records, and the precise time at which separate words were spoken could be seen from the second channel. The clarity of these records varied considerably from subject to subject. Although apparatus noise was low, there was interference due to brain rhythms and especially to impulses in the speech muscles (compare the silent and oral reading records). Thus a certain amount of visual filtering had to be done to discriminate between noise and eye movements. In fact many of the records had to be rejected because of the low degree of confidence with which the eye movements could be distinguished from interference.

(*a*)



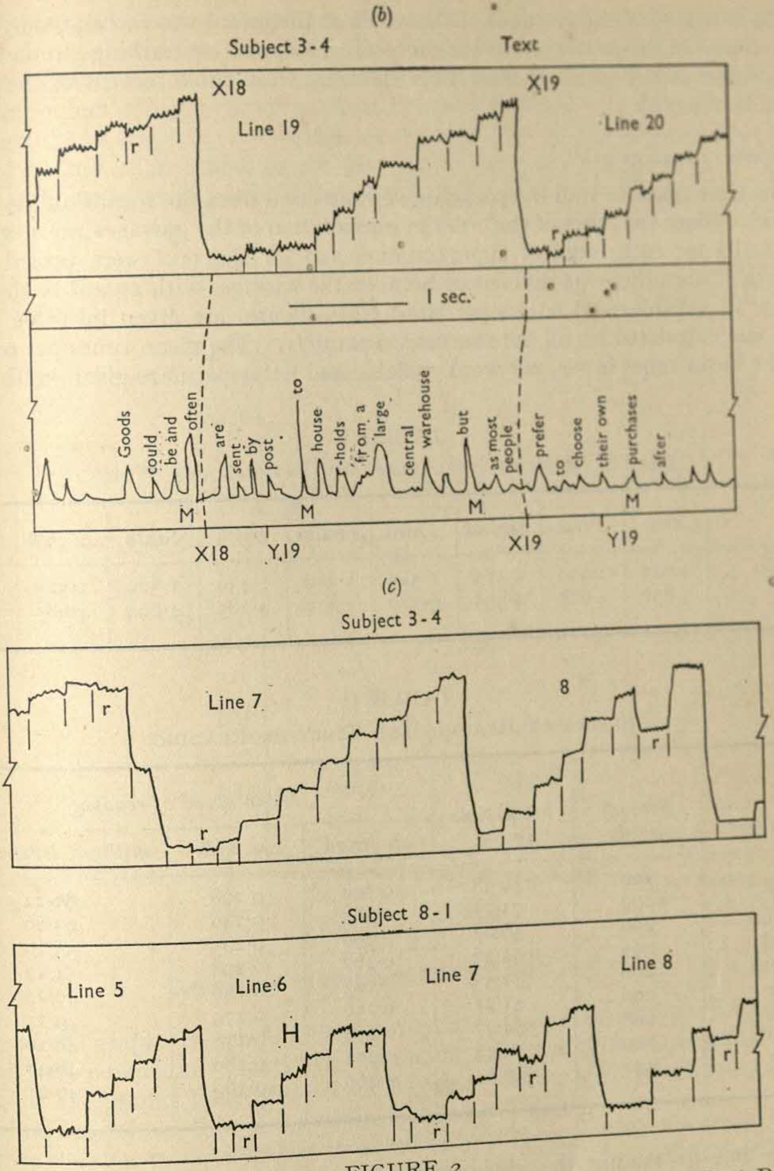


FIGURE 2

Sample records for eye movements. (*a*) and (*b*) Oral Reading; (*c*) Silent Reading.

*Notes*

1. r = regression. H = Head movement.
2. For both subjects there are fewer fixations per line in silent reading, and they are of shorter duration than those in oral reading.
3. A comparison of the records of subject 3–4 on silent and oral reading shows the interference due to the speech organs.
4. On the oral reading records, "M" represents a place marked on the record at the time the word was spoken.
5. "X" represents the time when the eye leaves the end of a line.
   "Y" is the time when the last word on the line has been spoken.
   "XY" is the eye-voice span.

The assigning of the positions of the words on the record was facilitated by noting the positions of the pauses on the tape recording, and also by marking (on the record) the position of key words as they were spoken. Only three records of the speech had to be rejected.

### 1. *Speed of reading*

The time taken to read the passages was measured from the record on the second channel. Since the effect of the order of presentation of the passages was negligible, the results for each order of approximation and for the text were pooled across subjects. Since there are differences between the passages with regard to the mean number of syllables and letters per word (these figures are given in Table I), the speed was calculated for all the measures separately. The mean times per passage, and the mean times in sec. per word, syllable and letterspace are given in Table II.

TABLE I

Mean Number of Syllables and Letters per Word

| | *Zero* | *First* | *Second* | *Third* | *Fourth* | *Fifth* | *Sixth* | *Eighth* | *Text* |
|---|---|---|---|---|---|---|---|---|---|
| *Syllables* .. | 2·615 | 1·490 | 1·415 | 1·345 | 1·410 | 1·430 | 1·520 | 1·515 | 1·540 |
| *Letters* .. | 7·850 | 4·675 | 4·335 | 4·280 | 4·410 | 4·285 | 4·695 | 4·685 | 4·707 |

TABLE II

Speeds of Reading and Time for Reading

| *Passages* | *No. of words* | *Mean time in sec.* | *Mean speed of reading* | | |
|---|---|---|---|---|---|
| | | | *sec./word* | *sec./syll.* | *millisec./letter-space* |
| Zero .. | 200 | 140·38 | 0·702 | 0·268 | 80·22 |
| First .. | 200 | 71·57 | 0·385 | 0·240 | 63·80 |
| Second .. | 200 | 58·07 | 0·290 | 0·205 | 55·20 |
| Third .. | 200 | 53·37 | 0·267 | 0·198 | 51·17 |
| Fourth .. | 200 | 52·53 | 0·263 | 0·186 | 49·23 |
| Fifth .. | 200 | 51·28 | 0·256 | 0·179 | 49·17 |
| Sixth .. | 200 | 54·07 | 0·270 | 0·178 | 48·10 |
| Eighth .. | 200 | 54·44 | 0·272 | 0·180 | 48·48 |
| Text .. | 444 | 115·54 | 0·260 | 0·169 | 46·21 |

The time for reading the whole passage decreases significantly only to the 3rd order, as in the Sumby and Pollack (1954) experiment. (All stated differences are significant at $p = 0{\cdot}05$ at least.) There is also a significant increase in time between the 5th and 6th order. This latter finding, which is unexpected from a simple informational viewpoint may be partly explained by considering the differences between the passages in the mean word length previously mentioned.

Time per unit is used as a metric since this measure is more normally distributed. For linguistic convenience the term "speed" will be used hereafter instead of its inverse. The mean speed of reading by sec. per syllable increases up to the 5th order, and the speed for the text is greater than that for any of the approximations. The results using sec. per letter-space are substantially the same.

*Differences between fast and slow readers.* The original hypothesis was that fast readers use context more effectively than slow readers and presumably use more context. Thus we would expect the fast readers to increase their reading speed up to higher orders of approximation than the slow readers. In order to test this prediction the 32 subjects were divided into two groups of 16 according to their mean speed of reading, as measured by sec. per syllable, over all the passages. The mean speeds of the two groups on the different passages are shown in Figure 3.

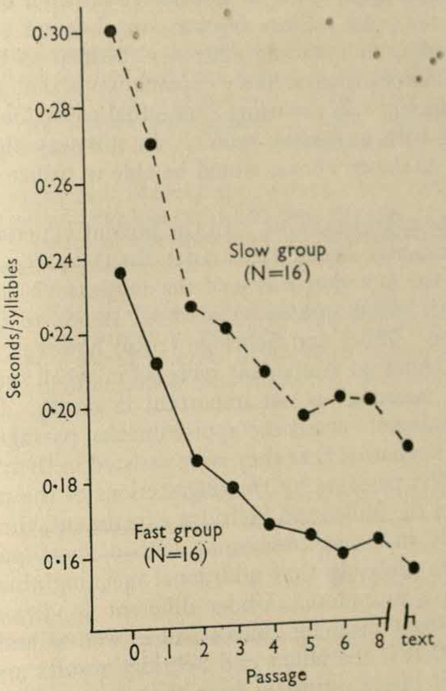


FIGURE 3

Comparison of speeds of fast and slow readers.

In all the graphs the passages are treated as if the successive orders of approximation were equidistant points in a continuum indicated by the abscissae, and the points on the graphs have been joined together. There is no justification for this procedure apart from that of convenience of presentation.

Since the slow group increases its speed only to the 5th order, we would predict a further increase in speed for the fast group between the 5th and 6th orders. A comparison of the mean changes in speed between the 5th and the 6th order approximations for fast and slow readers showed that the two groups differ; the fast readers increasing their speed significantly ($p < 0\cdot025$, Wilcoxon Test). There was no difference in speed between the 6th and 8th order passages for either group. Thus the faster readers may be said to be able to utilize the additional constraints found in the sixth order whereas the slow reader is limited in his use of contextual constraint to those found in the fifth order passage. This confirms the secondary hypothesis, and is compatible with the original hypothesis.

*Individual differences.* There was a wide range of reading speeds among the group. Thus, the times for reading the O-order passage ranged from 103 to 182 sec., and for the 8th order passage from 40 to 71 sec. Product-moment correlations for reading times were calculated between all pairs of passages and between all the passages and the mean time per subject. All the correlations were significant beyond the $p = 0{\cdot}001$ level, half of them being greater than $r = 0{\cdot}87$, and the lowest correlation, $r = 0{\cdot}727$, being between the times for the zero order passage and the text. The size of even this correlation might in part be due to differences in motivation or interpretation of instructions, but probably also indicates the presence of general oral reading skills which apply both to disconnected, often unfamiliar polysyllabic words and to connected prose. Since Morton (1959) found a relationship between amount of leisure reading and reading efficiency (defined as the product of (silent) reading speed and comprehension) a likely explanation is that people who read more, both develop more efficient ways of using contextual constraint, and also have more experience of dealing with unfamiliar words. In this way they would be superior on all passages, and, as shown above, would be able to utilize additional contextual constraints.

*"Meaning" and the approximations.* In the present experiment the performance on the text is significantly better ($p < 0{\cdot}01$) than that on the approximations. This was in spite of the fact that many of the subjects observed punctuation and meaning pauses, which added an average of about 12 sec. to the time of reading or 0·013 sec. per syllable. Miller and Selfridge (1950) however, in an experiment on immediate memory, found no equivalent increase in recall on their prose passages and concluded that "meaning" is not important in recall. This presupposes that there is nothing "meaningful" about the approximation passages. However, several of the present subjects reported that they were assisted in their reading of the fourth or fifth and higher order passages by treating sections of the passage as meaningful units. In addition, in the Miller and Selfridge experiment, the words were read out "slowly and distinctly in a near monotone with a short pause between words," (pp. 180–1). It seems probably that additional meaningfulness of the text would be obscured by such a procedure. Under different conditions of presentation it might well be found that "meaning" aids recall as well as assisting in reading. It might also be noted that if the Miller and Selfridge results are recalculated on the basis of a syllable unit (using estimates from the passages they quote on pp. 184–5 and the graph on p. 181) the amount of textual material recalled rises above that for the approximations.

### 2. *Errors*

The mean number of errors decreased to the 5th order. Assuming that errors are caused by excessive speed of reading, the differences in speed up to that point may be less than they would have been with error-free performance. There was no relationship between a subject's mean speed of reading and the number of uncorrected errors.

It might be noted that many errors were made and corrected, and no allowance was made for this waste of time in calculating the speed of reading. The effect of error-correction was to introduce a random element into the individual speeds, but it was assumed that this would average out for the group results. The total number of corrected errors was about the same for all passages.

The nature of the corrected and uncorrected errors is discussed elsewhere (Morton, 1961, 1964*b*).

### 3. *Eye-voice span*

The eye-voice span is a measure of the amount of material or time by which the voice lags behind the eyes in oral reading. Fairbanks (1937) has shown how the size of the material-span varies during reading. His results show that the span falls during regressions and refixations (i.e. a regression immediately following the first fixation on a line) and at the end of a sentence. Any pause in the speech which is unaccompanied by a regression resulted in the span lengthening, but at times, when a regression accompanied the correction of an error, the span would fall to nothing.

In order to measure the span, it is necessary to know precisely the fixation point of the eyes. In the present experiment, the position of the eyes could be told for certain only at the end of a line. On the record, the span on any line is the distance between the point where the eye begins the return sweep to the beginning of the next line and the point when the last word on the line had been spoken. This is illustrated on the sample records shown in Figure 2. The span was measured in two ways:

(*a*) *Material* (words, syllables, letter-spaces). These represent the amount of material which is no longer seen but has not yet been spoken at the instant the eye starts the return sweep.

(*b*) *Time*. This is the interval between the instant when the last word in the line was no longer being fixated and the instant it had been spoken.

As has been noted, a pause in the speech or a corrected error can alter the span considerably, and the relative effects of these upon time span and material span varied with the circumstances. Where such phenomena occurred within the measured span, that line was rejected if the pause was greater than 600 millisec., or if the error caused a regression back from the beginning of the next line. The levels of the material and time spans on the text would be reduced, relative to the other passages, by the tendency to observe meaning and punctuation pauses, and are so not strictly comparable.

*Results* (see Table III).

TABLE III

Mean Measures of Eye-Voice Span

| Passage | Time span in millisec. | Material span: In words | Material span: In syllables | Material span: In letter-spaces |
|---|---|---|---|---|
| Zero .. .. | 621 | 1·17 | 2·70 | 9·06 |
| First .. | 617 | 1·99 | 2·71 | 10·71 |
| Second .. | 618 | 1·99 | 3·34 | 11·25 |
| Third .. | 616 | 2·58 | 3·56 | 12·54 |
| Fourth .. | 636 | 2·93 | 3·30 | 13·04 |
| Fifth .. .. | 615 | 3·00 | 3·93 | 13·88 |
| Sixth .. | 673 | 2·61 | 4·23 | 13·86 |
| Eighth .. | 708 | 2·01 | 4·99 | 15·46 |
| Text .. .. | 641 | 2·78 | 4·14 | 15·27 |

The results, shown in this table and Figures 5–6, represent the means of measurements taken on the last five lines of legible record on each passage for each of 29 subjects. (Three of the speech records were undecipherable.)

(*a*) *Material span*. The mean number of words in the span increases to the 5th order and then falls. The drop appears to be related to differences in word length. When measured by syllables or letter-spaces, the material in the span increases, with some irregularities, up to the 8th order. This is true for both fast and slow readers (Fig. 4). Fast readers have the larger span, but the difference is not significant until the 6th order passage, by either syllables or letter-spaces. Thus the fast group appear to use that additional constraint between 5th and 6th order, not available to the slow group, which was deduced from their reading speeds.

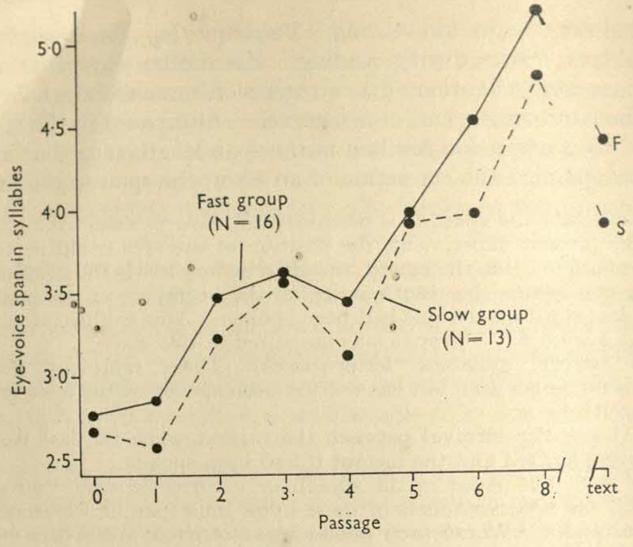


FIGURE 4

Eye-voice span in syllables for fast and slow readers.

(*b*) *Time span.* The time span remains approximately constant up to the 5th order and then increases to the 8th order. For all approximations the slow group has a larger mean time span (Fig. 5), but the difference is only significant for the 8th order ($p < 0{\cdot}05$, two-tailed: Mann-Whitney test). The average speed and the average time span for the approximations are related with $\tau = -0{\cdot}249$ ($p = 0{\cdot}026$).

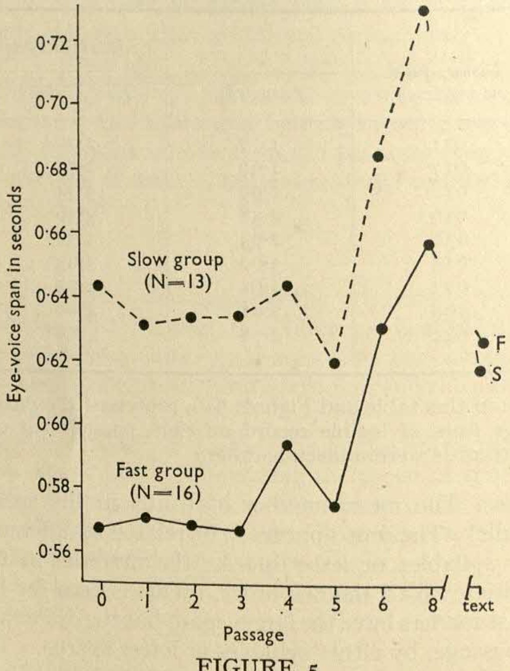


FIGURE 5

Eye-voice span in millisec. for fast and slow readers.

Fairbanks (1937) and Buswell (1920) also found that good readers had a higher material span than poor readers. However, in addition they reported a higher time span for good readers, a result not replicated in the present study. This might be due to differences in reading speeds between the studies.

*Dependence of time span upon material span and speed.* The two measures of material and time span are often treated as if they were in some way equivalent and independent of any other variable. Thus Fairbanks (1937) in referring to the concept of span remarks: "It is felt that 'lead' is more descriptive of the kind of eye-voice relation as it usually obtains," implying that 'lead' had the same significance in any unit; and Woodworth and Schlosberg (1954) say: "With experienced readers the span is as long as eight words, 2 seconds or a line of print, depending on which unit of measurement one prefers to consider" (p. 507).

However, it is clear that the rate of reading equals the material span divided by the time span (which we will write as $R = M/T$). At any speed of reading, either of the two measures can, theoretically, take any value, and the value of the other measure is then determined. With increasing constraint up to the 5th order, M rises roughly proportional to R while T remains roughly constant. It seems likely then that T is a variable dependent upon the values of R and M.

*Tendency to speed up at end of lines.* The equation $R = M/T$ can be used to estimate the speed of reading at the ends of the lines (where the span was measured). This speed measure was compared with the measured speed (S) for the passage. Apart from the 4th order, R is greater than S for both fast and slow readers. In general then, the reading speed is faster than average at the ends of lines. From Fairbank's (1937) result, that the material span at the ends of lines is approximately equal to the average for the passage, then the present measures of time span must be underestimates of the average value.

## 4. *Eye-movements*

Two eye-movement measures were taken on each of six lines for each passage, i.e. about half the record:

(*a*) *The mean duration of fixation.* For each line this was taken as the time between the beginning of the first fixation and the end of the last fixation on that line divided by the total number of fixations for that line. The measure included the time taken by the actual moves from one fixation point to the next, and accordingly all the times quoted are about 15 millisec. (Tinker, 1947) in excess of the true duration for a fixation.

(*b*) *The mean number of forward saccadic movements and regressive movements per line.* Sample records are presented in Figure 2.

The results are the average scores for the 18 subjects whose records could be interpreted with confidence. No more than six lines could be taken without reducing the number of subjects whose records could be included.

It should be noted that there is a variation in the eye movements from line to line in any one passage. For example, more regressive movements occur in those lines where reading errors are made and corrected. In addition the length of the saccadic movements and the duration of individual fixations have a wide range. The duration of individual fixations ranges from 180 to 350 millisec. on most passages for most subjects.

### *Results*

(*a*) *Duration of fixation* (fixation time). The mean fixation times for the different passages are given in Table IV. Each figure is the mean of from 80 to 120 fixations averaged over nine subjects for each of the sub-groups. The longer mean fixation time on the Zero order passage could be due to the fact that some eye movements on that passage were too small to be recorded; these might have occurred when a particularly unusual word was being carefully examined. Apart from this there appears to be no difference in fixation time among the various passages.

TABLE IV

Mean Duration of Fixation. Fast vs. Slow Readers
(measures in millisec.)

| *Passage* | *Fast group* ($n = 9$) | *Slow group* ($n = 9$) | *Difference* |
|---|---|---|---|
| Zero .. .. | 260 | 252 | 8 |
| First .. .. | 244 | 241 | 3 |
| Second .. .. | 238 | 243 | −5 |
| Third .. .. | 238 | 247 | −9 |
| Fourth .. .. | 234 | 242 | −8 |
| Fifth .. .. | 235 | 236 | −1 |
| Sixth .. .. | 243 | 235 | 8 |
| Eighth .. .. | 243 | 243 | 0 |
| Text .. .. | 245 | 238 | 7 |

An analysis of variance, from which the Zero order results were omitted, was carried out on the individual's scores, and is summarized in Table V. There are relatively large inter-subject differences and negligible intra-subject variations between passages. The "Group" source concerns the division between fast and slow readers referred to below. The individual mean fixation times ranged from 219 to 281 millisec.

TABLE V

Analysis of Variance of the Fixation Times

| *Source* | *D.F.* | *S.S.* | *M.S.* | *F* |
|---|---|---|---|---|
| Between groups (G) .. | 1 | 14·40 | 14·40 | < 1 |
| Between subjects .. | 16 | 32333·60 | 2020·80 | 15·432 ($p < 0·01$) |
| Between passages (P) .. | 7 | 891·22 | 127·32 | < 1 |
| P × G .. .. .. | 7 | 1290·60 | 184·37 | 1·408 |
| Residue .. .. .. | 112 | 14666·70 | 130·95 | |
| Total .. .. .. | 143 | 49196·52 | | |

The stability of the average fixation time has been noted previously by Buswell (1922), Robinson (1933) and Stone (1941) who measured eye-movements of subjects reading passages of varying difficulty. Whereas there were differences between the passages in the number of fixations, the duration remained constant. Tinker (1951) however, did find differences in mean fixation time between easy prose (220 millisec.) and scientific material (236 millisec.). Such a difference would be found if the eye made a number of undetectibly small fixations on unusual words, thus biasing the mean, as seemed to happen on the Zero order passage in the present experiment. Tinker gives no information which might help to settle that point.

The problem arises as to what determines the duration of a fixation. Tinker (1958) considers perception time, thinking time and reaction time to "eccentric stimulation" (p. 218) as being included in the total fixation time. However he regards 100 millisec. (an adequate tachistoscopic exposure "for a well cleared up perception"; p. 218) as suitable for the perception time, and quotes several figures for reaction time, centering on about 170 millisec. (the same order of magnitude found by the writer in his own rather crude check). The average fixation time in the present study is less than 250 millisec. for oral reading, and individual fixations, especially in silent reading, were as brief as 170 millisec. Thus the notion that eye-movements are triggered after analysis of the immediate stimulus is clearly inadequate as a general rule, since there is no time, or less than no time, available for either the initial clearing-up process, or the comprehension time. It often happens that an individual word will be fixated for an extended period, when the sequence described may take place, but in rapid reading some other mechanism clearly operates.

Carmichael (1957) goes even further than Tinker, and says: "... the measurement of fixation times after regressive movements must be related to complex brain activity and . . . it is associated with an identifiable *segment or unit of thought*. This technique demonstrates that this total *complex cognitive time* may be as short as one fifth of a second or less" (p. 7: present writer's italics). Carmichael seems to be making the mistake of supposing that all the comprehension processes relevant to an immediate stimulus must be completed before the eye can move on. There seems however to be no *a priori* reason to suppose that the visual and comprehension analyses (wherever and however they take place) could not function simultaneously on different material. Indeed the very existence of regressive movements indicates that the eyes move on before the material fixated has been adequately processed.

It seems then that the signal for the eyes to move can be given before the immediate stimulus has been analysed, and since the mean duration of regressions is the same as that of forward fixations, one need construct no special theories nor make special deductions regarding them.

(*b*) *Number of fixations and regressions*. The mean number of forward and regressive eye movements per line, and also the difference between these two measures, are given in Table VI. The total number of fixations decreased to the 5th order. This is a necessary concomitant of the invariance of the mean fixation time and the increase in speed to the fifth order. Of more interest is the small change in the difference measures. Up to two-thirds of the difference in the total number of forward movements could be contributed by the necessity of returning the eyes to the pre-regression fixation point and hence they are a direct result of the regression.

TABLE VI

MEAN NUMBER OF EYE MOVEMENTS PER LINE

| *Passage* | *Forward movements (F)** | *Regressive movements (M)*** | *Difference (F–M)**** |
|---|---|---|---|
| Zero . . . . | 16·31 | 6·04 | 10·27 |
| First . . . . | 14·58 | 4·58 | 10·00 |
| Second . . . . | 13·36 | 3·74 | 9·62 |
| Third . . . . | 12·22 | 3·36 | 8·86 |
| Fourth . . . . | 12·39 | 2·99 | 9·40 |
| Fifth . . . . | 10·77 | 2·88 | 7·89 |
| Sixth . . . . | 10·84 | 3·09 | 7·75 |
| Eighth . . . . | 11·38 | 2·89 | 8·49 |
| Text . . . . | 11·26 | 2·99 | 8·27 |

These differences are significant with a t-test $p < 0·025$ on the Wilcoxon test.

* $0 > 1 > 2 > 3, 4 > 5, 6, 8$, text

** $0 > 1 > 2 > 3 > 4, 5, 8$
$2 > 6$

*** $0, 1, 2, 4 > 5, 6, 8$, text
$1 > 3 > 5, 6$
$8 > 6$

In some cases the post-regression movement takes the eye beyond the pre-regression fixation point, but not always. It was not possible to obtain precise data from the present records.

Cherry (1957) discusses the possibility of applying the concept of "feed-back" to the reading situation, and wonders whether "the point to which the vision is shifted, in any one saccade, may be the result of a prediction based on previous perception" (p. 286). The implication is that more difficult passages would require shorter movements, and to some extent his idea is confirmed. However, the increase in the number of those regressive movements unconnected with errors as the order of approximation decreases indicates that the regulation of the saccade size is inadequate to compensate completely for the differences in difficulty of the passages.

In summary, it can be said that the mean duration of fixation in an oral reading task is independent of the redundancy of the passage and that the main effect upon eye movements is that there are more regressions in the less constrained passages, with the addition of a slightly smaller mean size of saccade.

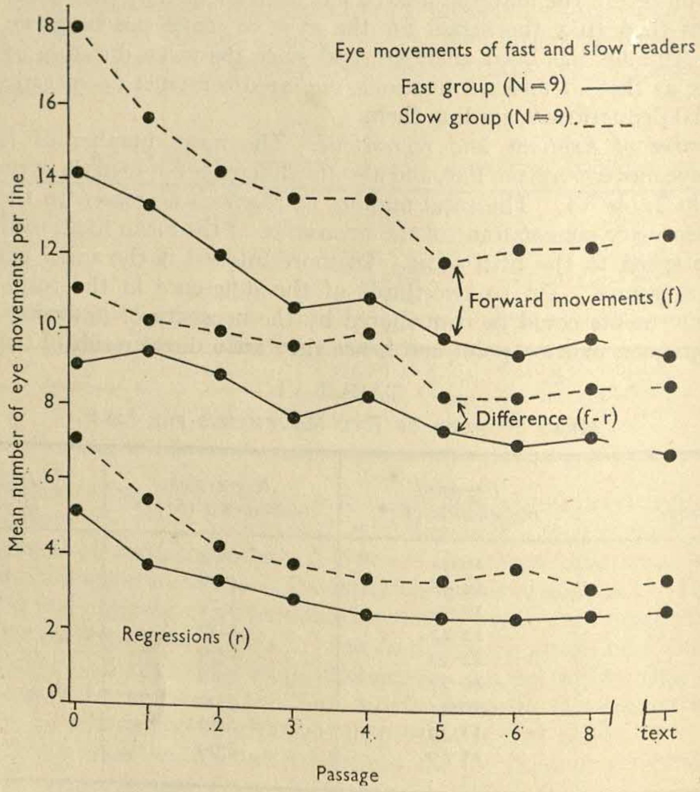


FIGURE 6

Eye movements for fast and slow readers.

The results are in agreement with those of Anderson and Swanson (1937), in as much as the superiority of the good readers in oral reading takes the form of fewer eye movements.

(*c*) *The correction of errors.* The eye-movement and speech records were examined to see if there was any relationship between the correction of an error and regressive movements. A corrected error, in this context, includes a word partially spoken correctly, and then repeated. Inadequacy of the records prevented examination of more than 33 corrections, a quarter of the total, but the results seem sufficient for certain general conclusions to be drawn.

(1) Only 23 of the 33 examples studied were accompanied by regressions. There was, furthermore, no relationship between the category (Omission, Substitution, Transposition, etc.) of an error and whether it had been accompanied by a regression or not.

(2) When a regression did occur, it occurred anywhere from 600 millisec. before the voice had paused after an error, to 600 millisec. after the pause. In other words, sometimes the regression enables the error to be noted, and sometimes a regression would be made because the error had been made and noted, perhaps via auditory feed-back.

(3) The correct word was spoken from between 250 and 600 millisec. after the start of the regression fixation.

## Conclusions

The amount of contextual constraint in a passage affects directly the speed of reading, the amount of material in the eye-voice span and the mean number of fixations and regressions per line. Comparisons of changes in reading speed and the size of the span show that fast readers can utilize more contextual constraint than slow readers. The eye-voice span measured in time units can be shown to be dependent upon the material in the span and the speed of reading. The mean duration of fixation remains unchanged on different passages.

The research reported here was done while the author was in receipt of a D.S.I.R. studentship at the Department of Psychology, Reading University. I wish to express my gratitude to Professor M. D. Vernon who supervised the work. I also wish to acknowledge the help of Dr. E. F. R. W. Crossman at all stages in the work and of various of my present colleagues who have discussed many of the problems.

## References

Anderson, I. H., and Swanson, D. E. (1937). Common factors in silent and oral reading. *Psychol. Monogr.*, **48**, 61–9.

Buswell, G. T. (1920). An experimental study of the eye-voice span in reading. *Suppl. educ. Monogr.*, No. 17.

Buswell, G. T. (1922). Fundamental reading habits: a study of their development. *Suppl. educ. Monogr.*, No. 21.

Carmichael, L. (1957). Comprehension time, cybernetics and regressive eye movements in reading. Paper presented at the Fifteenth International Congress of Psychology, Brussels.

Cherry, C. (1957). *On Human Communication.* New York.

Fairbanks, G. (1937). The relationship between eye movements and voice in oral reading of good and bad silent readers. *Psychol. Monogr.*, **48**, 78–107.

Miller, G. A., and Selfridge, J. A. (1950). Verbal context and recall of meaningful material. *Amer. J. Psychol.*, **63**, 176–85.

Morton, J. (1959). An investigation into the effects of a Reading Efficiency Course. *Occup. Psychol.*, **33**, 222–38.

Morton, J. (1961). *Reading, Context, and the Perception of Words.* Unpublished Ph.D. thesis: Reading University.

Morton, J. (1964*a*). The effect of context on the visual duration threshold for words. *Brit. J. Psychol.*, **55**, 165–80.

MORTON, J. (1964*b*). A model for continuous language behaviour. *Lang. Speech*, 7, 40–70.

ROBINSON, F. P. (1933). Role of eye movements in reading with an evaluation of the techniques for their improvement. *Iowa St. Univ. Aims of Res.*, **39**.

SHACKEL, B. (1958). A rubber suction cup surface electrode with high electrical stability. *J. appl. Psychol.*, **13**, 153–8.

SHACKEL, B. (1960). Pilot study in electro-oculography. *Brit. J. Ophthal.*, **44**, 89–113.

SHACKEL, B., SLOAN, R. C., and WARR, N. J. J. (1958). Detector plots eye movements. *Electronics*, **31**, 36–39.

SHANNON, C. E. (1951). Prediction and entropy of printed English. *Bell syst. Tech. J.*, **30**, 50–64.

STONE, L. G. (1941). Reading reactions for various types of subject matter: an analytical study of the eye movements of college freshmen. *J. exp. Educ.*, **10**, 64–77.

SUMBY, W. H., and POLLACK, I. (1954). Short-term processing of information. HFORL Report. TR-54-6.

TAYLOR, A., and MORAY, N. (1960). Statistical approximations to English and French. *Language and Speech*, **3**, 7–10.

TINKER, M. A. (1947). Time relations for eye movement measures in reading. *J. educ. Psychol.*, **38**, 1–10.

TINKER, M. A. (1951). Fixation pause duration in reading. *J. educ. Res.*, **44**, 471–9.

TINKER, M. A. (1958). Recent studies of eye movements in reading. *Psychol. Bull.*, **55**, 215–31.

WOODWORTH, R. S., and SCHLOSBERG, H. (1954). *Experimental Psychology*. New York.

*Manuscript received 25th October*, 1963.

# SHORTER ARTICLES AND NOTES

## LEFT-RIGHT DIFFERENCES IN THE PERCEPTION OF MELODIES

BY

DOREEN KIMURA

*From the Montreal Neurological Institute*

Normal subjects were given two auditory tests, one consisting of spoken digits presented dichotically, the other of melodies presented dichotically. On the Digits test, the score for the right ear was higher than for the left (as previously established), and on the Melodies test the score for the left ear was higher than for the right. These findings were related to the different roles of the right and left hemispheres of the brain in verbal and nonverbal perception.

### INTRODUCTION

The neural substrate for musical ability has been the subject of much inconclusive discussion. The traditional view, based on clinical studies of "amusia," is that the left hemisphere, particularly the left temporal lobe, is the area most concerned with music functions (Henschen and Schaller, 1925; Ustvedt, 1937). Most of these studies suffer from the twofold disadvantage of an incomplete knowledge of the lesion, and an inadequate testing of musical function nearly always contaminated by verbal factors in instruction, naming, etc. A recent study free of these drawbacks found instead that efficient performance on certain nonverbal auditory tasks rests more with the right temporal lobe than with the left. Milner (1962) reported that scores on the Timbre and Tonal Memory subtests of the Seashore Measures of Musical talents were depressed by right temporal lobectomy but not by left temporal lobectomy.

It has previously been shown, in normal subjects, that the predominant role of the left hemisphere in the perception of speech sounds is reflected in better recognition of verbal material arriving at the contralateral ear (Kimura, 1961*a*). When different spoken digits were presented simultaneously to the two ears, those digits arriving at the right ear were more efficiently recognized than digits arriving at the left. This situation is possible because the crossed auditory pathways are more effective than the uncrossed (see Kimura, 1961*b*). One might therefore expect that the predominance of one hemisphere for the perception of musical sounds should reveal itself in an asymmetry analogous to that for speech sounds. Specifically, if the right temporal lobe is more involved in musical recognition than the left, it should be possible to demonstrate a left-ear superiority for musical sounds, in the same subjects in whom the right ear is more efficient for verbal material.

### METHOD

Two auditory tasks were presented on different days. They will be described in detail below. For the tests a dual-channel tape-recorder and earphones were employed. The volume for each ear was set to a comfortable standard level, and this level was checked before every test session. For half the subjects on each test the earphones were reversed so that any remaining asymmetries in the tape or apparatus were counterbalanced over ears. Throughout these tests the subject was discouraged from selectively attending to one ear.

#### *Subjects*

Subjects were 20 female student nurses and postgraduate nurses. They were all right-handed.

#### *Digits test*

This test, employing variations of the Broadbent technique (Broadbent, 1954), was identical with the test used in the earlier study (Kimura, 1961*a*). It consisted of the presentation of groups of spoken digits in pairs so that different numbers arrived at the two ears simultaneously. After three such pairs (six digits), the subject reported all the numbers he had heard, in any order he liked. Maximum score for each ear is 96.

*Melodies test*

The individual melodies were excerpts from solo passages in concertos by Mozart, Telemann, Vivaldi, Bach, and Antonini. Most of the instruments recorded were woodwinds, but there were occasional violin, viola, and cello excerpts. Eighty such passages were tape recorded from commercial records and subsequently classified into 20 sets of four. Within each set of four melodies, the same instrument was used, and an attempt was made to have the pitch range and the tempo very similar, leaving as the main clue the melodic pattern. The original passages were then re-recorded to make melodies of 4-sec. duration. For each set, two of the four melodies were first played simultaneously on the two separate channels, so that one melody was heard in one ear at the same time that the other was played in the other ear. There was then a 4-sec. silence and the four melodies were played in succession, in normal binaural fashion, i.e. with the same melody in each ear. There was a 3-sec. gap between melodies. Thus the two melodies first heard dichotically were repeated separately, and the subject had simply to identify which two they were. He would reply, for example, "They were the second and third," or "First and fourth," whatever he thought was correct. The position of the repeated melodies was varied in a counterbalanced manner from set to set. The score for each ear was the number of correct identifications for that ear. Since the first two were used for practice, there were 18 sets in the test proper, making a total possible score of 18 for each ear.

## RESULTS

Results are shown in Table I. On the Melodies test, the score for the left ear is significantly superior to that for the right, whereas on the Digits test, the right ear is superior (*t* test for correlated means). From Table II it can be seen that 16 of the 20 subjects have a higher score for the left ear on the Melodies test, and 15 of the same 20 subjects have a higher score for the right ear on the Digits test. There is thus a clear-cut differentiation between relative ear superiority on the two types of material.

TABLE I

MEAN TOTAL CORRECT (N = 20)

| | *Left ear* | *Right ear* | t *ratio* |
|---|---|---|---|
| Melodies .. .. | 13·55 (75 per cent.) | 11·3 (63 per cent.) | 3·57** |
| Digits .. .. | 86·7 (90 per cent.) | 90·3 (94 per cent.) | 2·83* |

** $p < 0.01$
* $p < 0.02$

TABLE II

NUMBER OF SUBJECTS SHOWING DESIGNATED EAR SUPERIORITY ON TWO TESTS (N = 20)

| | *Better on left* | *No difference* | *Better on right* |
|---|---|---|---|
| Melodies .. .. | 16 | 2 | 2 |
| Digits .. .. | 3 | 2 | 15 |

## DISCUSSION

The fact that melodies presented to the left ear are more accurately recognized than those arriving at the right supports Milner's (1962) view that the right temporal lobe plays a greater role in nonverbal auditory perception than does the left. Due to the greater effectiveness of the crossed pathways, melodies arriving at the left ear are more efficiently transmitted to the right temporal lobe, an area most important for their perception, than are melodies arriving at the right ear. An analogous but opposite effect occurs for verbal material presented to the two ears. Thus the left-right differences which occur in the present study reflect an asymmetry of function in the two cerebral hemispheres.

The asymmetry is not due to a difference in the sensitivity of the two temporal lobes to the frequency characteristics of sound, for the pure tone loss after temporal lobectomy is very slight, and the pattern of loss is the same after right temporal lobectomy as after left (Sinha, 1959). Rather, the differentiation appears to be along the verbal-nonverbal dimension, and among nonverbal sounds music may be especially effective in eliciting a left-ear effect. This is suggested by a preliminary study employing the presentation of different numbers of clicks to the two ears simultaneously. The results again showed a trend for the left ear to have a higher score, but, unlike the results for melodies, the effect was not significant. The right hemisphere may thus be especially important for the perception of melodic patterns. It does not follow, however, that it is dominant for all musical functions, for both Milner's study (1962) and the present one have emphasized the perceptual aspects of musical ability.

The fact that performance on the Seashore tests is not affected by the addition of Heschl's gyrus to a standard temporal lobectomy (Milner, 1962) suggests that the difference in auditory function between right and left temporal lobes does not depend on a difference between the primary auditory areas. It appears, for musical ability at least, that the observed left-right asymmetries depend on some areas of brain beyond Heschl's gyrus. Although the efferent connections of the primary auditory area in man are not completely known, there appear to be direct projections to the surrounding superior temporal gyrus (Bonin, 1950) and the posterior insular gyrus (Pfeifer, 1920) on the same side. It seems reasonable to consider these direct pathways as responsible for the differences found in the present study, since areas more remote from Heschl's gyrus should have a higher probability of being stimulated equally by ipsilateral and contralateral ears.

The asymmetries observed here occur only under conditions of dichotic stimulation. In an unpublished study, the Timbre test of the Seashore battery was presented to a group of normal subjects, one ear at a time, on two separate occasions. This procedure yielded no difference between ears. Similarly, the right-ear effect for digits occurs to a significant extent only with dichotic presentation. One reason for this may be that dichotic listening puts more demands on the system than does monaural listening. However, there is probably another factor involved. Rosenzweig (1951) has suggested that the auditory system is so arranged that some central units in each half of the brain fire to stimulation of the ipsilateral ear, some to the contralateral ear, and some to both. More units are activated by contralateral stimulation than by ipsilateral, but in addition, in those units which fire to both, the contralateral connections occlude the ipsilateral. Thus the greater effectiveness of the contralateral pathways should become more apparent when both ears are stimulated, but with different material. Let us consider what this means in terms of the role of the right temporal lobe in the recognition of melodies. When different melodies are presented to the two ears, there are two possible pathways to the right temporal lobe: the ipsilateral pathway (from the right ear) and the contralateral pathway (from the left ear). [We shall ignore for the present argument the pathways to the left temporal lobe and subsequent commisural connections to the right, since this would be a much slower connection.] When only one ear is stimulated at a time, the difference between ears may not be great enough to permit detection of a difference. When different information travels along these pathways, however, those units which fire to both ears will be taken up by the contralateral (in this case, left ear) pathway, according to Rosenzweig's system. In this way, dichotic stimulation may enhance the difference between the two pathways.

An assumption basic to this argument is that the main difference between ipsilateral and contralateral pathways is in the number of units they command. Results obtained from animal studies by both Tunturi (1946) and Rosenzweig (1951) suggest that this is so. They found that if a click was presented to the ear, and records taken from ipsilateral and contralateral auditory cortex, the only difference between the responses at these two sides was in amplitude. There was no detectable difference in either latency or duration. Thus it does appear that it is not the earlier arrival of impulses from the contralateral pathways which results in the asymmetries found in this study, but rather, it is due to the greater number of impulses.

This research was supported by Grant B-2831 to Dr. Brenda Milner from the U.S. Public Health Service. I am indebted to Dr. Peter Milner for advice and assistance in preparing the tapes, and to the nurses of the Montreal Neurological Institute who acted as subjects.

## REFERENCES

BONIN, G. v. (1950). *Essay on the Cerebral Cortex.* Springfield.

BROADBENT, D. E. (1954). The role of auditory localization in attention and memory span. *J. exp. Psychol.*, **47**, 191–6.

HENSCHEN, S. E., and SCHALLER, W. F. (1925). Clinical and anatomical contributions on brain pathology. *Arch. Neurol. Psychiat.*, **13**, 226–49.

KIMURA, D. (1961*a*). Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psychol.*, **15**, 166–71.

KIMURA, D. (1961*b*). Some effects of temporal-lobe damage on auditory perception. *Canad. J. Psychol.*, **15**, 156–65.

MILNER, B. (1962). Laterality effects in audition. In MOUNTCASTLE, V. (Ed.), *Interhemispheric Relations and Cerebral Dominance.* Baltimore.

PFEIFER, R. A. (1920). Myelogenetisch-anatomische Untersuchungen über das kortikale Ende der Hörleitung. *Abh. math.-phys. Kl. sächs. Akad. Wiss.*, **37**, No. 2.

ROSENZWEIG, M. R. (1951). Representations of the two ears at the auditory cortex. *Amer. J. Physiol.*, **167**, 147–58.

SINHA, S. P. (1959). *The Role of the Temporal Lobe in Hearing.* Unpublished Master's thesis. McGill Univ.

TUNTURI, A. R. (1946). A study on the pathway from the medial geniculate body to the acoustic cortex in the dog. *Amer. J. Physiol.*, **147**, 311–9.

USTVEDT, H. J. (1937). Ueber die Untersuchung der musikalischen Functionen bei Patienten mit Gehirnleiden, besonders bei Patienten mit Aphasie. *Acta med. Scand.*, Supp. 86.

*Manuscript received* 15*th July*, 1963

# ACCURACY OF RECOGNITION FOR SPEECH PRESENTED TO THE RIGHT AND LEFT EARS

BY

D. E. BROADBENT and MARGARET GREGORY

*From the Medical Research Council Applied Psychology Research Unit, Cambridge*

Earlier reports by other authors had shown that recall of dichotically presented speech was better for words arriving at the right ear, while recognition of passages of music was better for music presented to the left ear. It seemed desirable to test recognition for speech, in order to show that the difference between materials rather than the difference in testing techniques was responsible for this disparity between ears.

Accordingly 18 men were presented with dichotic lists of digits, and then tested for recognition of the digits. There was no indication of an advantage for the left ear, but rather of the advantage for the right already established for recall of speech materials.

## INTRODUCTION

It is now well established (Bryden, 1963; Kimura, 1961) that when two sets of spoken digits are presented simultaneously to the two ears of a listener, he will be more likely to recall successfully those on the right ear as opposed to those on the left. Kimura (1964) has, however, shown that if passages of music are presented simultaneously in the same way, it is the passage on the left ear which is most likely to be recognized. Because of the difficulty of requiring the listeners themselves to reproduce the musical stimulus, recognition of the correct item out of several alternatives was necessarily employed by Kimura. It might conceivably have been argued therefore that the distinction in this case is not the use of music rather than speech as the stimulus, but rather the use of recognition rather than recall as the testing technique. The present experiment used speech stimuli and recognition as the method of testing.

## METHOD

### *Apparatus and material*

The material for 24 trials was presented on a Ferrograph twin channel recorder through headphones, one channel to each ear. Following the usual convention, we will call stimulation identical on the two ears "binaural" and that which differs on the two ears "dichotic." Each trial consisted of the number of the trial, presented binaurally, followed by three pairs of digits arriving dichotically and simultaneously at the rate of one pair per ½ sec. These pairs of digits were followed, after an interval of 1 sec., by four groups of three digits presented binaurally at the rate of one digit per ⅓ sec. and with a ½ sec. gap between each group. Of the four binaural triads presented for recognition, one triad corresponded to the digits previously given to the right ear, and one triad corresponded to that given previously to the left ear. The subject had to pick out these two triads. For example:

| | | | | | | |
|---|---|---|---|---|---|---|
| R ear only | | 483 | | | | |
| Both ears | Trial 1 | | 483 | 403 | 039 | 038 |
| L ear only | | 039 | | | | |

The gains of the two channels were adjusted so that as far as possible the dichotic lists sounded approximately the same loudness in both ears and the binaurally presented material was heard as more or less in the centre of the head. If the subject found these conditions far from being fulfilled the gains were adjusted accordingly.

The material was composed so that in each trial the two dichotically presented triads had one digit in common which was not in the same serial position. The two binaural triads which were not identical with either of the dichotic lists had, however, two items in common with one of them and one with the other. This was found to be essential to avoid recognition on a basis of partial perception. In each trial one of the correct triads was in the first two alternatives of the recognition test, and one was in the second pair.

Each of the four possible arrangements of the two correct triads occurred with equal frequency, and the first correct alternative was in equal numbers of cases that from the right ear and from the left.

### *Subjects and procedure*

Eighteen naval ratings between the ages of 18 and 30 were tested individually, each having one presentation of the 24 trials. Nine of the subjects had the material recorded on the upper track of the tape presented to the right ear and that on the lower to the left, and the other nine vice versa. This was to rule out any advantage that might be attached to lists recorded on one particular channel and also to allow for the possibility that the particular lists recorded on one of the tracks might be easier to remember than those on the other.

Each subject was given an explanation with the aid of diagrammatic examples of what he would hear. He was instructed to write down after each trial which two of the four groups for recognition corresponded to the lists he had heard separately on the two ears. After the test was over the subject was asked whether he was right- or left-handed.

The two scores taken were the number of errors made on the right ear and on the left. That is, for each ear the number of trials in which the correct alternative corresponding to that ear had not been identified.

## RESULTS

The results showed a tendency for more errors to be made on the lists presented to the left ear. The percentages of wrong identification for the two ears are given below.

| R ear | L ear |
|---|---|
| 22·9 per cent. | 26·9 per cent. |

Thirteen subjects made more errors on the lists presented to the left ear than on those arriving on the right, and five subjects made more errors on the right than on the left. This result is just significant $p = 0{\cdot}05$ by a one-tailed test.

There were three subjects who said they were left-handed and wrote with their left hands; all three made more errors on the lists presented to the left ear.

## DISCUSSION

While the results would not perhaps by themselves be very convincing evidence of a priority for the right ear, they are sufficient to confirm the difference already established by Bryden and Kimura. There is no indication that recognition as opposed to recall reverses the advantage of the right ear. Thus Kimura (1964) must be taken as demonstrating that the use of music rather than speech gives an advantage to the left ear.

Thanks are due to Dr. Brenda Milner for the discussion which gave rise to this experiment.

## REFERENCES

BRYDEN, M. P. (1963). Ear preference in auditory perception. *J. exp. Psychol.*, **65**, 103–5.

KIMURA, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psychol.*, **15**, 166–71.

KIMURA, D. (1964). Left-right differences in the perception of melodies. *Quart. J. exp. Psychol.*, **16**, 355–8.

*Manuscript received 2nd October, 1963*

# REPETITION AND THE RECOGNITION OF PAIRED ASSOCIATE ITEMS

BY

PETER B. WARR

*From the Department of Psychology, University of Sheffield*

It is suggested that incremental learning of paired associate items will be observable if learning is measured by recognition rather than by recall. An experiment is reported which fails to support this hypothesis. Some factors which may be influencing the performance of subjects are briefly discussed.

The writer has recently reported the results of a series of experiments investigating the effect of repeated presentation of items on learning in varying conditions (Warr, 1963, 1964). These experiments have consistently failed to provide evidence for the hypothesis that learning proceeds on a wholly incremental basis. It is however possible that wholly incremental learning may be revealed if use is made of a recognition measure. The somewhat restricted experimental conditions in which verbal learning is usually studied may in fact conceal some learning, which might become apparent if learning were to be assessed by recognition rather than by recall. For instance, subjects who are unable to recall a response in the 3 sec. provided might exhibit learning of that response if they are asked to recognize it from an ensemble of several possible responses.

The present experiment is a direct replication of an earlier one (Warr, 1963), with the exception that learning is measured by recognition of responses instead of by recall. It was predicted that the results would support the incremental hypothesis, since a recognition measure is presumably more sensitive than the customary assessment of learning by recall score.

The experimental procedure involved six presentations of paired associate lists. Each presentation (of 12 items) was followed by a recognition test, during which each stimulus term was presented with five possible response terms, one of which was correct. The subjects were aware that the correct response was among these terms. As in the previous experiment, the stimuli were letters or double letters, and the responses were drawn haphazardly from the population of two-digit numbers. (No numbers including a zero and no double numbers were employed.) Subjects were free to make no response to a test item. Experimental conditions, time intervals and instructions were the same as those used in the earlier experiment. Twenty-three subjects from the same population as previously, took part in the experiment.

The effect of repetition upon learning in these conditions is shown in Table I. It can be seen that the probability of correct response, "p(C)," assessed here by recognition tests, declines with increasing numbers of successive presentations of items. This apparently

TABLE I

PROBABILITY OF CORRECT RESPONSE (RECOGNITION) AFTER EACH NUMBER OF PRESENTATIONS

| | *Number of successive presentations* | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Items presented | 783 | 417 | 234 | 125 | 64 | 33 |
| Items correct | 297 | 154 | 78 | 39 | 19 | 9 |
| p(C) .. .. | 0·379 | 0·369 | 0·333 | 0·312 | 0·297 | 0·273 |

indicates that repetition hinders learning in these conditions rather than improving it. There is however a valid objection to this conclusion. The decline in p(C) with increasing numbers of presentations is an artefact, since some subjects do not require six presentations to learn all the items. In other words, not all subjects are contributing to the later values of p(C). This artefact can be removed by analysing the results for the individual subjects in terms of the rank correlation between p $(C_1)$, $p(C_2) \dots p(C_n)$ and number of presentations. When this analysis was carried out the mean value of *rho* was found to be +0·01. A significant positive correlation of this kind indicates that repetition has resulted in improved learning; in this case no benefit has been derived by the group as a whole from repeated presentation of items. It is of interest, however, that individual differences in the effect of repetition are marked. The distribution of *rho's* (to the nearest 0·1) is shown in Table II.

TABLE II

DISTRIBUTION OF RANK CORRELATIONS BETWEEN $p(C_1)$, $p(C_2) \dots p(C_n)$, AND NUMBER OF PRESENTATIONS FOR THE 23 SUBJECTS

| | *Frequency* | |
|---|---|---|
| *rho* | *Positive* | *Negative* |
| 0·0 | 0 | 0 |
| 0·1 | 1 | 0 |
| 0·2 | 1 | 2 |
| 0·3 | 1 | 1 |
| 0·4 | 2 | 3 |
| 0·5 | 1 | 1 |
| 0·6 | 0 | 3 |
| 0·7 | 2 | 1 |
| 0·8 | 1 | 0 |
| 0·9 | 1 | 0 |
| 1·0 | 1 | 1 |

Further analysis of the data suggests a possible interpretation of the variance exhibited in this Table. Those subjects for whom a negative correlation was obtained had learned a mean of 22·33 items by the end of the experiment, whereas those subjects for whom the correlation was positive had learned a mean of 29·82 items. This difference is significant ($U = 18$; $p < 0{\cdot}01$). It is possible, therefore, that faster learners benefit from repetition, but that slower learners do not. This hypothesis is of course open to more direct test.

The finding that the group as a whole has not benefited from repetition of items might be partly explained in terms of the persistent emission of the same incorrect responses. The use of a fixed ensemble of available responses in a recognition test allows calculation of the frequency of repetition of incorrect responses expected by chance; such a calculation involves many uncertain assumptions when a recall measure is used. A $\chi_2$ comparison of the observed frequency of repetition of incorrect responses with the expected frequency (from an expansion of the multinomial distribution) indicated that errors were persisting (despite correction on intervening presentations) at a level well above chance. On the other hand, the observed frequency only exceeded the expected frequency for items not learned after three presentations. For items learned by the third presentation observed and expected frequencies were not significantly different. Since the majority of items were learned by the third presentation (see Table I), persistence of errors cannot be wholly responsible for the failure of subjects in these conditions to benefit from repetition. It is however likely that this is at least partly responsible for the steady decline in p(C) values for the slower learners with increasing numbers of presentations.

It may be noted that in the original experiment, in which the procedure was identical to the present one in all respects except that a recall measure was employed, $p(C_{2+})$ was significantly higher than $p(C_1)$. The failure to find such a difference here is somewhat surprising; as mentioned above it cannot be explained in terms of persistence of errors. Neither is an explanation in terms of the more general concept of "item difficulty" plausible, since such a factor would presumably also be involved in the original results. The findings are probably best interpreted in terms of a genuine difference between recognition and recall, evidence for which has been obtained in many studies. Further work is required to provide a higher-order explanation of this difference.

This work was supported by a grant from D.S.I.R. to the Department of Psychology, University of Sheffield. The writer is grateful to Miss H. M. Davies, Department of Statistics, University of Sheffield for her advice on statistical treatment of the results.

## REFERENCES

WARR, P. B. (1963). The effect of repeated presentation of items on paired associate learning. *Quart. J. exp. Psychcl.*, **15**, 262–72.

WARR, P. B. (1964). The effect of repetition of items on two phaces of paired associate learning. *Quart. J. exp. Psychol.*, **16**, 224–31.

*Manuscript received 3rd April, 1963.*

# IMMEDIATE MEMORY AND THE "PERCEPTION" OF LETTER SEQUENCES

BY

A. D. BADDELEY

*From the Medical Research Council Applied Psychology Research Unit, Cambridge*

It is suggested that the relationship observed by Miller, Bruner and Postman (1954) between the redundancy and the accuracy of reproduction of tachistoscopically presented letter sequences is not a perceptual effect as they suggest, but is due to the informational limitations of immediate memory. An experiment is performed which shows an exactly similar relationship between redundancy and number of letters correctly reproduced when exposure time is long enough for the subject to read out, and hence perceive, all the material with complete accuracy. It is concluded that the more redundant the letter sequence and the longer the exposure time, the more effectively the sequence can be coded and the better it will be recalled.

## INTRODUCTION

The experiment to be described is concerned with a study by Miller, Bruner and Postman (1954) on the reproduction of tachistoscopically presented letter sequences. The sequences used, each comprising eight letters, formed zero, first, second or fourth-order approximations to English (Shannon, 1948). These sequences were flashed on a screen in random order at each of the six exposure times ranging from 10 to 500 millisec. The number of letters correctly reported was found to depend on both duration of exposure and type of material: the longer the exposure and the higher the order of approximation, the better the performance. When the various approximations were corrected for redundancy, it was found that the amount of information reproduced at each exposure was the same for all four approximations.

This result has been interpreted by Miller *et al.*, and by subsequent writers (Cherry, 1957, p. 280–4; Garner, 1962, p. 252–5; Vernon, 1962, p. 109–10) as a *perceptual* effect. An attempt to use these data to obtain an estimate of the time required to perceive such letter sequences accurately, however, suggests that such an interpretation is unjustified. This is shown quite clearly if one tries to estimate for how long an eight-letter random sequence would need to be exposed in order to be perceived correctly. At 500 millisec., the maximum exposure time used, subjects were only reporting about 50 per cent. of each random sequence correctly, and it is therefore necessary to extrapolate. Fortunately, the relationship between information intake and duration of exposure is relatively clear (see Miller *et al.*, 1954, or Cherry, 1957, p. 283). When plotted against log exposure time, information intake gives a negatively accelerated curve. A conservative extrapolation yields a "perception" time for the 37·6 bits of information contained in a random sequence of eight letters of between one and two *hours*. If, however, the subject "perceives" the eight letters as two groups of four, his "perception" time is reduced to approximately 1 sec., while if instead he "perceives" the item as four pairs of letters, this should require a total of only 150 millisec. While it is unwise to place too much faith on the details of results based on extrapolation, those obtained in this case are so implausible that it is difficult to avoid the conclusion that either the function describing the rate of intake of information by the visual system must change completely at some point beyond the 500 millisec. exposure duration, or else that some factor other than perception time must have influenced this result. A closer examination of the procedure used suggests the latter.

The experiment required the subjects first to perceive the letter sequences, then to write down what they saw. This obviously involves remembering what was seen, and it seems likely that a sequence of eight random letters would be beyond the span of immediate memory. Thus, even if the subject did perceive all eight random letters, his score would be limited by his memory span. This conclusion is supported by the work of Sperling (1960), who demonstrated an immediate memory span of about 4·5 tachistoscopically presented random letters, which in contrast to the results of Miller *et al.*, was not influenced by exposure time over a range of 15 to 500 millisec.

It seems possible then, that the result obtained by Miller *et al.*, may reflect ease of remembering as a function of order of approximation to English, rather than the rate at which the visual system can take in information. The following experiment examines this hypothesis by presenting the material used by Miller *et al.*, at a rate which is slow enough for the subject to read out eight random letters, and requiring him to write down what he saw. Under these conditions, where the subject clearly has long enough to perceive the letter sequence, any errors that occur might reasonably be attributed to forgetting rather than to his perceptual limitations. It was predicted that the relationship between order of approximation to English and the percentage of letters correctly reported would parallel that shown by Miller *et al.*

## METHOD

### *Materials*

The material comprised the 60 eight-letter sequences used by Miller *et al.* (1954), which consisted of 15 zero-order, 15 first-order, 15 second-order and 15 fourth-order approximations to English. An additional series of eight-letter sequences was prepared for use in the pretest. Each sequence was typed in block capitals on a 4½ × 3½ in. white card.

### *Procedure*

The material was presented manually, being placed in front of the subject for the appropriate exposure time, then covered over with a blank card. Presentation time was controlled by synchronizing the experimenter's movements with a metronome beating at the appropriate rate.

Twelve naval ratings and six research workers acted as subjects. Both groups were given a pretest which involved reading out a number of random sequences of eight letters. In the case of the naval ratings, each item was exposed for 2 sec. For all but two subjects, this proved long enough to read the items, the criterion being five successive correct responses. For the two slower readers, the presentation time was increased until they reached the criterion at a presentation time of 2·2 sec. per item. In the case of the six research workers, the presentation time was determined for each subject by the method of limits, again using a criterion of five successive correct responses. The resulting presentation times ranged from 1·1 to 1·3 sec. per item with a mean of 1·2 sec. The exposure time used in the main test was that at which the subject had reached criterion in the pretest.

The main experiment involved presenting the 60 items used by Miller *et al.*, and requiring the subject to write down on a response sheet as much of each as he could remember. The stimulus cards were divided at random into three packs of 20, with the constraint that each pack should contain five items of each approximation. Within each pack items occurred in a predetermined random order. The order in which the three packs were presented was counterbalanced so that each of the six possible orders occurred equally often.

## RESULTS

The number of correctly located letters was scored for each response and was converted into a percentage. This is the "placement" score used by Miller *et al.*, for their main analysis. Figure 1 shows the relationship between this score and the per cent. redundancy of the four orders of approximation for both groups of subjects, together with the results obtained by Miller *et al.*, at their maximum exposure time of 500 millisec. There is a very clear relationship between order of approximation to English and amount recalled in both naval and research subjects. The two groups were combined in an over-all analysis which showed that the effect of approximation to English was highly significant across both subjects and items. Thus, subjects found the fourth-order easier than the second, which was easier than the first, which in turn was easier than the zero-order material ($p < 0{\cdot}001$, Sign Test in each case). Similarly, a comparison across items using Jonckheere's test (Jonckheere, 1954) showed that they tended to be recalled in the predicted order of difficulty ($z = 6{\cdot}67$, $p < 0{\cdot}001$). Finally, it is clear from Figure 1 that the slope of the line relating redundancy and percentage correct for both groups in the present study is very similar to that obtained by Miller *et al.*, and interpreted as due to a perceptual effect.

It has been shown by Baddeley (1963, ch. IV) that the ease of learning a pair of nonsense syllables can be estimated on the basis of the "predictability" of the syllables, that is, on how closely their letter structure resembles that of English. A predictability score was therefore computed for each of the eight-letter sequences used in the present study as follows. A statistical "subject" is assumed, who attempts to guess each letter of the sequence, given the previous letter, on the basis of the relative frequencies of digrams in the English language. The mean number of "guesses" required per letter is the measure of predictability. When the predictability score for the 60 eight-letter items was correlated with recall score, it yielded a rho of 0·80, showing a highly significant relationship ($t = 10{\cdot}15$, $d.f. = 58$, $p < 0{\cdot}001$). It is interesting that the correlation should be so high despite the fact that predictability, being based on the digram structure of English, is incapable of discriminating between the second and fourth-order approximations, which differ only when sequences of three or more letters are considered.

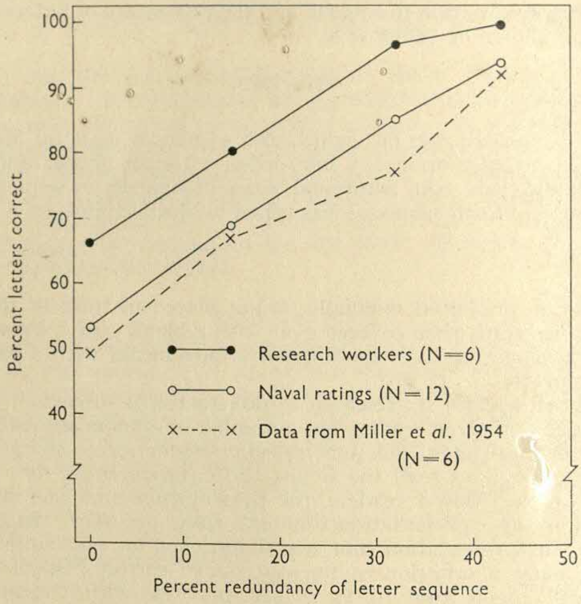


FIGURE 1

Accuracy of reproducing letter sequences as a function of their redundancy.

The mean amount of information produced was computed for the four levels of approximation, correcting each on the basis of its redundancy (Miller *et al.*, 1954).* The mean information scores for zero, first, second and fourth-order approximations were 21·61, 22·91, 23·83 and 21·37 bits per sequence respectively. A Friedman test suggested that consistent differences occurred between approximations ($p < 0{\cdot}01$, two tail). The six possible paired-comparisons were made using a Wilcoxon test. These showed only one significant difference, that between the second and fourth-order approximations ($p < 0{\cdot}01$, two tail). This is probably because the fourth-order approximations contain only 22 bits of information, so that even perfect performance would not convey as much information as subjects are able to reproduce with first and second-order approximations. Closer examination of individual scores suggests that the better subjects were being constrained by the amount of information available, even in the first and second-order approximations.

---

* As Tulving (1963) has pointed out, Shannon's estimate of the redundancy of printed English assumes an infinitely long sequence of letters and thus should not be applied as by Miller *et al.*, directly to eight-letter sequences. In recalculating the redundancy scores, however, Tulving neglects Shannon's use of the space as a letter. Taking this into account, the estimate used by Miller *et al.* is accurate for all but the fourth-order approximation, which should have a mean redundancy of 40·5 per cent. rather than 43 per cent.

The overall picture suggests a tendency to reproduce more information as approximation to English increases, a tendency which is limited by the fact that less information is available at the higher order approximations. If such a trend exists, it should appear when longer letter sequences, containing more information, are used.

## DISCUSSION

Miller *et al.*, have shown a relationship between order of approximation to English and the accuracy with which tachistoscopically presented letter sequences are reported. Their study, however, required the subject not only to perceive, but also to remember each letter sequence. The present study ensures that sufficient time is available for the sequence to be perceived and yet produces an almost identical relationship between redundancy and performance. It therefore seems likely that most, if not all, the effect they attribute to the limited capacity of the visual system is in fact due to limitations in immediate memory.

An interesting discrepancy occurs between the results of Miller *et al.*, which show an increase in the number of letters reported as exposure time lengthens, and the observation of Sperling (1960) that number of letters reported remains constant over the range of exposure times used by Miller *et al.* However, Sperling showed that this limitation is set by the subjects' immediate memory span, and when the memory load was reduced by a sampling technique which required the subject to report only part of the stimulus array, the visual system proved capable of handling at least 40 bits of information. It therefore seems likely that increased exposure time improved performance in the study by Miller *et al.*, not because it allowed the subject to perceive more letters, but rather that it gave him more time for coding the eight letters into a smaller number of "chunks" and thus increased the memory span (Miller, 1956). The more closely a letter sequence approximates to English the easier it will be to encode in terms of existing language habits, and thus the better it will be remembered.

## REFERENCES

BADDELEY, A. D. (1963). *Some Psychological Aspects of the Coding of Information.* Unpublished Ph.D. dissertation, Cambridge University.

CHERRY, C. (1957). *On Human Communication.* New York.

GARNER, W. R. (1962). *Uncertainty and Structure as Psychological Concepts.* New York.

JONCKHEERE, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, **41**, 133–45.

MILLER, G. A. (1956). The magical number seven, plus or minus one. *Psychol. Rev.*, **63**, 81–97.

MILLER, G. A., BRUNER, J. S., and POSTMAN, L. (1954). Familiarity of letter sequences and tachistoscopic identification. *J. gen. Psychol.*, **50**, 129–39.

SHANNON, C. E. (1948). A mathematical theory of communication. *Bell. Syst. Tech. J.*, **27**, 379–423.

SPERLING, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.*, **74**, No. 11.

TULVING, E. (1963). Familiarity of letter sequences and tachistoscopic identification. *Amer. J. Psychol.*, **76**, 143–6.

VERNON, M. D. (1962). *The Psychology of Perception.* London.

*Manuscript received 28th July, 1963.*

# SELECTIVE STRATEGIES IN THE ASSIMILATION OF SUCCESSIVELY PRESENTED SIGNALS

BY

A. F. SANDERS

*From the Institute for Perception RVO-TNO, Soesterberg, The Netherlands.*

It is found that performance in experiments on the psychological refractory period is highly affected by instructions. In the present experiment subjects were instructed either to handle the signals successively or to group them. Both instructions were obeyed. This seems to indicate that the human organism has various strategies available. Which strategy is actually applied is likely to depend on the structure of the experimental situation, and partly also, on momentary preference. This may explain the variety of results in the literature on this subject.

Finally, it is found that, when two signals are presented simultaneously, the total reaction time is considerably shorter if the signals are "grouped" than if they are "handled successively." This difference disappears at interstimulus intervals of 0·2 sec. and 0·4 sec. The hypothesis is put forward, that the gain in time at the former case is due to simultaneous perceptual processing of the signals.

It has been found in several studies during the last decade that, when two signals are presented simultaneously or in rapid succession in order to be responded to, the reaction times are delayed in comparison to the reaction time when only one signal is given.

The phenomenon is clearly established as far as the occurrence of the delays is concerned. Less consensus is present, however, about the actual values of the delays. In fact they tend to vary with the structure of the experimental task.

In some situations the picture is clear cut. Here the relation could be described by:

$$RT_2 = 2\ RT_1 - I\ (I \leqslant RT_1) \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

where $RT_1$ and $RT_2$ represent reaction time to the first and second signal respectively, and where I stands for the intersignal interval. This relation has been found in tasks where simple reactions are asked in combination with ungraded responses (Davis, 1956, 1957, 1959). Using graded responses, the formula had to be changed slightly, but the general idea—i.e. that the signals are handled successively—remained unaffected (Welford, 1952, 1959).

The picture is quite different, however, if choice reactions are asked for. In that case the only clear phenomena are, first, that $RT_2$ shows some delay at least and, secondly, that $RT_1$ is also delayed—at least in most cases (Davis, 1962; Borger, 1963; Sanders, 1963). Especially the latter phenomenon was absent in the original experiments, albeit not completely. Welford (1952, 1959) reported that at very small intervals ($I \leqslant$ 0·1 sec.) both $RT_1$ and $RT_2$ were delayed, and similar results were obtained by Elithorn and Lawrence (1955) and Halliday, Kerr and Elithorn (1960).

Welford has introduced the term "grouping" to indicate this effect, suggesting that sometimes the signals are transmitted simultaneously. This is also assumed by Broadbent (1958), who argues that, if the intake of information occurs discontinuously in successive perceptual samples, it can be expected that signals are sometimes transmitted in the same sample. This will be especially the case when they are presented simultaneously or in rapid succession, since it is likely that the duration of perceptual samples is highly limited. We may depict this process by:

$$RT_1 = RT_2 + I - c\ (I < RT_1) \quad \ldots \quad \ldots \quad \ldots \quad (2)$$

Where c stands for the motor time to carry out the second response after the first one. It has been argued that, in the case of grouping, the reactions must be made simultaneously (Halliday, Kerr and Elithorn, 1960; Borger, 1963). This does not seem a necessary condition, however, since grouping has to be considered foremost as a central affair. The actual reactions may be carried out in succession (Welford, 1959).

Now, it is not likely that the results in choice reaction studies can be satisfactorily explained in terms of grouping. First, we expect this process at small intervals only, and secondly, the experimental data only partly fit formula 2. For the rest, values are obtained that vary between those predicted by formulæ 1 and 2. Starting from the idea that the human organism acts like a single channel system in the transmission of information—and this model still seems always more adequate than a multichannel hypothesis, in view of the fact that delays in reaction time have been found everywhere—it is necessary to explain this variation in terms of a successive handling process.

Now, successive handling is not necessarily bound to formula 1. It is possible that at the arrival of the second signal ($S_2$), the processing of the first signal ($S_1$) is momentarily interrupted in order to store the newly arrived signal. This may occur especially if it is known when $S_2$ is to be presented as when the interval between $S_1$ and $S_2$ is constant. Another idea is that the response to $S_1$ is postponed until $S_2$ has been transmitted centrally (Davis, 1962; Borger, 1963; Sanders, 1963). Both hypotheses require the assumption of storing mechanisms, apart from that at the entrance of the transmission channel (Broadbent, 1958).

In this way we can discern several transmission strategies, with grouping and successive handling according to formula 1 as extremes. Subjects may be capable of switching from one to another strategy, although under some conditions one special strategy may be predominant—as in the simple reaction studies of Davis and Welford.

If this idea is valid, we must expect that the results in experiments on refractoriness are highly dependent on the instructions how to react to the signals. Subjects can be instructed either "to complete $RT_1$ without taking any notice of $S_2$," or "to collect all perceptual data before any response is carried out."

With the former instruction we should obtain data that fit formula 1, while in the latter case formula 2 is expected to give a satisfactory description. In fact, Hick (1948) has considered the possibility that the transmission of successive signals would be affected by instructions, but the issue is neglected in more recent studies.

Formula 2 is supposed to be found at both small and larger intervals, albeit grouping is only expected at the small ones and at simultaneous presentation, while at larger intervals a composite successive handling process is assumed. However, if $RT_1$ is postponed until $S_2$ has been transmitted centrally, it is likely that the actual picture of the reaction times will be very similar to that when grouping occurs.

We wonder if it is possible to differentiate between these processes. If not, the theoretical value of grouping becomes very dubious.

Differentiation may be possible, however, by means of the consideration that, in the case of grouping, the transmission process should be more efficient. Two items of information are processed at once, while in any successive handling operation one item is transmitted every time. From the results on the relation between information and processing time, we may infer that the former process runs quicker, although it is admitted that this argument is not quite watertight. Taking this for granted, however, we can expect that, when subjects are instructed "to group" or "to handle successively" in successive experimental series, the total reaction time will be significantly shorter if the signals are really grouped. Such an effect will not be found if, instead of grouping, a "composite successive handling" process is followed, since there are no clear arguments for the assumption that the different types of successive handling will differentiate in efficiency. This reasoning follows especially if the signals are presented simultaneously. When an interval between $S_1$ and $S_2$ is present, the grouping process supposes that the transmission of $S_1$ is postponed until the arrival of $S_2$—and this will tend to cancel out the gain in time, obtained by the simultaneous processing (see also Welford, 1959). The longer the interval, the more efficient will "composite successive handling" become. In this way an experiment on the effect of instruction on performance in refractoriness, may contribute to the theoretical status of "grouping." An experiment was therefore designed to test the following hypotheses:

(1) Instructions will have a large effect on the reaction times to two signals which are presented simultaneously or in rapid succession.

(2) At small intervals the instructions to group will produce a shorter total reaction time than the instructions to handle the signals successively. This difference will disappear at larger intervals.

### METHOD AND PROCEDURE

In a semi-dark room the subject was seated at a table on an adjustable chair, at a distance of 70 cm. from a screen. $S_1$ and $S_2$ were projected slightly above eyelevel to the left and right from the subject's meridian. The distance between $S_1$ and $S_2$ was 7·5 cm.

Both $S_1$ and $S_2$ consisted of a column of bright dots, the number being either four or five per column. The diameter of the dots and the distance between two successive dots were both 1 cm. approximately. The alternative signals had an equal chance of occurrence. The task was to carry out two choice reactions, one to $S_1$ (four of five dots) and another to $S_2$ (idem). Four morse keys, corresponding to the various possible signals, were fixed on the table, behind which the subject was seated and they served as reaction set. The responses to $S_1$ were given with the left hand and those to $S_2$ with the right hand. Three seconds before the arrival of $S_1$ a visual warning signal was presented in the place where $S_1$ would appear.

The apparatus is schematically drawn in Figure 1. The experimenter started the apparatus, which was immediately followed by the presentation of the warning signal (w) from the projection unit. The presentation took place by opening the shutter, placed before the appropriate projector. At the same time $S_1$ and $S_2$ were selected by the stimulus selector and after a time delay of 3 sec. the selected shutters of both $S_1$ and $S_2$ were opened. Thus $S_1$ became visible on the screen, while $S_2$ was held back a moment, until a second shutter was opened, having a time delay of $(3 + x)$ sec. from the start. Reaction times were measured by means of two time counters, that were started at the presentation of $S_1$ and were stopped after the response to $S_1$ and $S_2$ respectively. The correctness of the responses was checked by a set of lightbulbs, indicating which signals and which responses had been given.

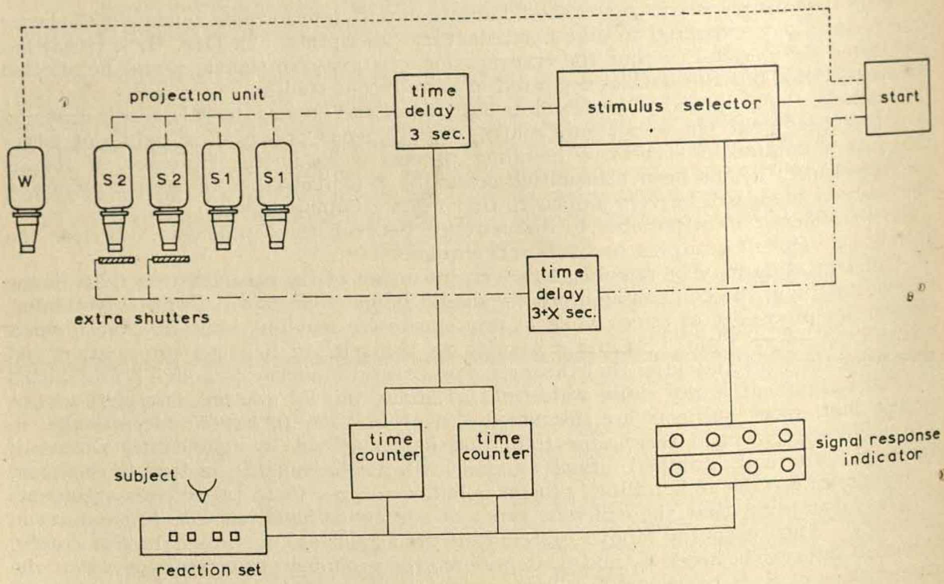


FIGURE 1

Much care was taken to avoid any cues from the clicks of the shutters. In a preparatory study, it was checked whether the subjects could predict from the clicks which signal was given. The results showed that this was not the case. Using the same experimental set up, an experiment was carried out on the effect of choice reactions on the delays of $RT_1$ and $RT_2$ (Sanders, 1963), which produced results, similar to those obtained by Davis (1962) and Borger (1963). This may serve as an argument that the present experimental situation has no special features from which diverging results are to be expected.

Eighteen subjects—naval ratings—were randomly divided in three equal groups. Each group was tested twice at one of three intervals between $S_1$ and $S_2$: 0·0 sec., 0·2 sec. and 0·4 sec. In one trial the subjects were instructed "to complete $R_1$ before taking notice of $S_2$," and in another trial they were asked "to collect all perceptual data before preparing any response." The order of the trials was counterbalanced.

Before the experiment started, an intensive training series was given on both strategies. Each subject was trained only at that interval that was given to him in the experiment. In an experimental session 20 pairs of correct reactions were obtained. Accuracy was stressed in the instructions and the few erroneous reactions were omitted from treatment and replaced by correct reactions. The median values of the series were calculated separately for each subject.

### RESULTS

The means of the median reaction times per subject are presented in Table I. Moreover the total reaction time ($RT_{tot} = RT_2 + I$) is included. To check whether the subjects obeyed the instructions, predictions according to formulae 1 and 2 are also given.

TABLE I

MEANS OF MEDIAN REACTION TIMES IN SECONDS AT DIFFERENT INTERSIGNAL INTERVALS AND UNDER DIFFERENT INSTRUCTIONS

| *Instruction* | *Interval* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0·0 sec. | | | 0·2 sec. | | | 0·4 sec. | | |
| "*Successive handling*" | $RT_1$ | $RT_2$ | $RT_{tot}$ | $RT_1$ | $RT_2$ | $RT_{tot}$ | $RT_1$ | $RT_2$ | $RT_{tot}$ |
| Obtained .. .. | 0·46 | 0·87 | 0·87 | 0·37 | 0·54 | 0·74 | 0·44 | 0·48 | 0·88 |
| Predicted .. .. | | 0·92 | | | 0·54 | | | 0·48 | |
| "*Grouping*" .. | | | | | | | | | |
| Obtained .. .. | 0·65 | 0·72 | 0·72 | 0·60 | 0·48 | 0·68 | 0·89 | 0·53 | 0·93 |
| Predicted .. .. | | 0·65–0·80 | | | 0·40–0·55 | | | 0·49–0·64 | |

The predictions are derived from formulae 1 and 2. The factor $c$ in formula 2 is unknown but can vary probably between 0·00 and 0·15 sec. (Sanders, 1963).

### STATISTICAL ANALYSIS

To decide upon the differences in efficiency between "successive handling" and "grouping," three analyses of variance were carried out on the $RT_{tot}$, one for each group. A significant effect was obtained at the interval 0·0 sec. ($F = 14·35$, $p < 0·01$); while such evidence was absent at the remaining intervals ($F = 2·15$ and $2·47$ respectively, $p > 0·05$).

The obtained values are generally equal to those that are predicted, except at 0·0 sec. at the instruction "successive handling," where a mean divergence of 0·05 sec. appears. The significance of this divergence was tested by calculation of the 5 per cent. confidence limits of the mean of $RT_2$. The latter limit proved to be 0·039 sec. so that the obtained value is probably shorter than expected.

### DISCUSSION

The results appear to confirm hypothesis 1: The relation between $RT_1$ and $RT_2$ proves to be very much dependent on instructions. Except for successive handling with simultaneous presentation, the signals are processed exactly in the way the subject is asked to do. Now, in the present experiment two strategies were employed which gave opposite results. Presumably intermediate values between "successive handling" and "grouping" could be obtained if the subjects were asked "to interrupt the processing of $S_1$ momentarily at the arrival of $S_2$ in order to store the latter." Nevertheless, it is not likely that the process is only determined by instructions.

In that case no exclusive successive handling would have appeared in the simple reaction experiments of Davis. In those situations, successive handling may be predominant, since the latter type of reaction "closely resembles a reflex and is in effect a temporary prepared reflex" (Woodworth and Schlosberg, 1954). Presumably the $S - R$ process is less easily interrupted in simple reactions, than in choice reactions which clearly exceed the reflex level.

Another point of interest is the relation between the length of the interval and the selective strategy. The present results indicate that the instruction "successive handling" was obeyed beautifully at intervals of 0·2 and 0·4 sec., but not completely at simultaneous presentation. Apparently the subjects were not able to ignore $S_2$ completely in the latter case, although the reaction times were largely in the expected direction. As suggested in the second hypothesis, grouping was more efficient at simultaneous presentation, while no significant differences were found at intervals of 0·2 and 0·4 sec. We may conclude from this finding that the subjects were grouping the signals in the strict sense at the former condition only. In this way, differentiation between "grouping" and "composed successive handling" seems possible. The mean time gained at simultaneous presentation amounted to 0·15 sec. Assuming that this can be completely ascribed to the advantage of perceptual grouping, we can conclude that at an interval of 0·2 sec., this gain will be more than cancelled out already. Grouping is likely, therefore, to be restricted to intervals that are smaller than 0·15 sec., which corresponds with the existing suggestions from the literature (Welford, 1952, 1959; Sanders, 1963).

As a final conclusion from the present experiment, we can say that future research on refractoriness should be engaged with the question of which selective processes are preferred and which not under various experimental conditions, rather than with an endless search for one "unequivocal process" that can be applied everywhere.

I am indebted to Dr. J. A. Leonard and Dr. E. C. Poulton of the A.P.R.U., Cambridge, for reading and criticizing this paper.

## References

Borger, R. (1963). The refractory period and serial choice reactions. *Quart. J. exp. Psychol.*, **15**, 1–12.

Broadbent, D. E. (1958). *Perception and Communication.* Oxford.

Davis, R. (1956). The limits of the "psychological refractory period." *Quart. J. exp. Psychol.*, **8**, 24–38.

Davis, R. (1957). The human operator as a single-channel information system. *Quart. J. exp. Psychol.*, **9**, 119–29.

Davis, R. (1959). The role of "attention" in the psychological refractory period. *Quart. J. exp. Psychol.*, **11**, 211–20.

Davis, R. (1962). Choice reaction times and the theory of intermittency in human performance. *Quart. J. exp. Psychol.*, **14**, 157–67.

Elithorn, A., and Lawrence, C. (1955). Central inhibition: some refractory observations. *Quart. J. exp. Psychol.*, **7**, 116–27.

Hick, W. E. (1948). Discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, **1**, 36–51.

Halliday, A. M., Kerr, M., and Elithorn, A. (1960). Grouping of stimuli and apparent exceptions to the psychological refractory period. *Quart. J. exp. Psychol.*, **12**, 72–89.

Sanders, A. F. (1963). The selective process in the functional visual field. Ph.D. Thesis, Utrecht University.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high-speed performance—a review and a theory. *Brit. J. Psychol.*, **43**, 2–19.

Welford, A. T. (1959). Evidence of a single channel decision mechanism, limiting performance in a serial reaction task. *Quart. J. exp. Psychol.*, **11**, 193–210.

Woodworth, R. S., and Schlosberg, H. (1954). *Experimental Psychology.* New York.

*Manuscript received 5th July, 1963.*

# INDUCTION OF A PRINCIPLE

BY

CARL P. DUNCAN

*From Department of Psychology, Northwestern University, Illinois*

Two experiments are reported in which student subjects attempted to discover a principle obtaining among pairs of numbers and letters. In the first experiment, subjects were more successful when they were free to select whatever number-letter pairs they wished than if they were restricted in whole or in part to pairs specified by the experimenter. In the second experiment, subjects who did discover the principle were compared to those who did not. Successful subjects were shown to be slightly more systematic in their approach to the task, to work at a faster pace, to write down more positive instances, and to have a much stronger tendency to vary only one variable of the task at a time.

Wason (1960) reported a study in which subjects tried to discover a principle applying to instances each of which consisted of three numbers. He discovered that subjects mostly showed enumerative rather than eliminative induction, i.e. they made little use of instances that would have enabled them to eliminate wrong hypotheses. Wetherick (1962) found that by modifying Wason's procedure in certain ways, enumerative behaviour was reduced and the subject's chance of discovering the principle was increased. Elimination of hypotheses remained infrequent.

In the Wason study, the subject was free to write down, for each instance, any three numbers he wished, i.e. his choice of instances was not restricted in any way. This may account in part for the behaviour Wason observed. In the first experiment to be reported here, the subject is restricted, in various degrees, in his choice of instances, to determine whether the principle is discovered more (or less) readily under such restriction.

## EXPERIMENT I

*Task.* The task consisted of pairing the numbers 1–8 with the 10 letters A–J. The principle was that any of the four odd numbers could be paired with any of the last five letters (F–J), and any of the even numbers could be paired with any of the first five letters (A–E). Thus, of the 80 possible number-letter pairs, 40 were instances of the principle, the other 40 were not.

The task was presented on a sheet of paper at the top left of which the eight numbers were dittoed in normal order in a horizontal row. At the top right were the 10 letters in a row in normal order. The rest of the sheet contained 80 blanks, numbered consecutively from 1–80, in four columns. There were two forms of this sheet. In one, the blanks contained neither numbers or letters. In the other, each blank already had printed in it one of the eight numbers. Each number appeared once in each block of eight blanks. The order of assignment of numbers to blanks was random within each block.

*Conditions.* There were four conditions, representing different degrees or kinds of restriction in the subject's choice of numbers and letters to create pairs. (A pair always consisted of a number followed by a letter.) A different group of subjects worked under each condition. In Group U (unrestricted), the subject, using the sheet with nothing in the blanks, chose both the number and the letter to write in each blank. In Group RL (restricted letters), the subject, using the same sheet as Group U, chose numbers to write in the blanks, but the experimenter told the subject what letter to write in after each number. In Group RN (restricted numbers), the subject used the data sheet with numbers already entered in the blanks, but was free to choose letters. In Group RNL (restricted numbers and letters), the subject also used the sheet with the numbers in blanks, and the experimenter told the subject what letter to write after each number.

*Subjects.* The subjects were students from classes in introductory psychology. They were required to serve in experiments as part of the course work. Students were assigned to the four conditions in turn. There were 50 subjects in each of the four groups.

*Procedure.* The experimenter gave each subject a pencil and the sheet needed for the condition to which the subject had been assigned. Then the experimenter pointed to the numbers and letters at the top of the sheet and said, "See these numbers and letters. They have been paired according to a certain principle. I want you to try to discover what that principle is. This is the way we will do it." At this point, the experimenter inserted instructions (see below) appropriate for each group. Then the experimenter concluded, "As soon as you have any idea as to what the principle is, tell me. If you are wrong, or only partly right, I will tell you so, and we will continue with pairs until you want to make another guess as to what the principle is. Try to discover the principle as soon as possible. As soon as you discover the principle, the experiment is finished."

In Group U, the subject was told to write any number and any letter in a blank and to state the number and letter aloud. Then the experimenter would tell him whether the pair thus created was or was not an instance of the principle. In all groups, the subject was told to make a check mark after pairs that were instances of the principle, to help them remember.

In Group RL, the subject was told to write any number in a blank, and state the number aloud. Then the experimenter told the subject what letter to write after the number, and when the subject had done so, told the subject whether the pair was an instance of the principle. The experimenter had a key which specified what letter to give the subject for every pair, and whether the pair was or was not to be an instance of the principle. Whether a pair was to be a correct instance or not had been decided by flipping a coin, with the restriction that within a block of 40 pairs, half would be correct, half incorrect. In so far as possible, the experimenter used each letter once, in giving letters to the subject, within each block of 10 pairs.

In Group RN, the experimenter pointed out to the subject that a number, selected from those listed at the top of the sheet, had already been printed in each blank. The subject was told to write any of the letters after the number, and state both number and letter aloud. Then the experimenter told the subject if the pair was an instance of the principle.

In Group RNL, the experimenter also pointed out to the subject that there was a number in each blank, and told the subject to say the number aloud. Then the experimenter told the subject what letter to put after the number, and when the subject had done so, told the subject if the pair was an instance of the principle. The key used for giving letters to the subject, and determining whether each pair was to be a correct instance or not, was the same as that for Group RL.

The subject was allowed to work at his own pace. He could guess at the principle at any time and as often as he wished. He had to state the principle completely to be considered correct. Statement of the particular numbers and letters that could be paired, without, e.g. using phrases such as "odd numbers," was allowed. If the subject had not stated the principle correctly at the end of 25 min. of work, he was stopped and was recorded as a nonsolver.

## Results and Discussion

Two measures reflecting attainment of the principle are shown in Table I. One is the number of solvers, those subjects in each group that stated the principle correctly. The other is the median number of pairs (of numbers and letters) to solution (scores for nonsolvers were considered to be infinity). The results are quite similar with both measures. Numerically, the best performance is by the least restricted group (U), next best by the most restricted group (RNL). Poorest, and essentially equivalent, performance was by the two partially restricted groups (RL and RN).

TABLE I

Number of Subjects Who Discovered the Principle and Median Number of Pairs to Solution in Experiment I

| *Group* | *N* | *Solvers* | *Median pairs* |
|---|---|---|---|
| U | 50 | 43 | 39·5 |
| RL | 50 | 26 | 92·0 |
| RN | 50 | 27 | 82·0 |
| RNL | 50 | 37 | 60·0 |

Statistical analysis reveals the following: In number of solvers, Group U was significantly different from Group RL ($\chi^2 = 11·97$, $p < 0·001$, 1 *d.f.*), and from Group RN ($\chi^2 = 10·71$, $p < 0·01$), though not from Group RNL ($\chi^2 = 1·56$). Group RNL had significantly more solvers than Group RL ($\chi^2 = 4·29$, $p < 0·05$). Other comparisons were not significant.

Each group was compared to every other group by a median test on median pairs to solution. Group U differed significantly from Groups RL ($\chi^2 = 10·24$, $p < 0·01$), RN ($\chi^2 = 12·96$, $p < 0·01$), and RNL ($\chi^2 = 9·00$, $p < 0·05$). No other comparison was significant.

Examination of the number of incorrect (including partly correct) attempts to state the principle revealed no clear differences among the four conditions. The mean number of such attempts was 1·3, 0·7, 1·3, and 1·3 for Group U, RL, RN, and RNL, respectively. These means are based on all 50 subjects (solvers and nonsolvers) in each group. As might be expected, nonsolvers made somewhat more incorrect guesses (overall mean = 1·5) than solvers (mean = 1·0), since all guesses of nonsolvers were incorrect. However, it is clear that nonsolvers did not fail simply because they hesitated to make overt guesses and so failed to get information.

The results of Experiment I seem fairly clear. More subjects induced the principle, and in fewer pairs, when they were left free to create any pairs they wished. When a subject is unrestricted, he has the advantage that he can work on any hypothesis he chooses to entertain, and can switch at any time from one hypothesis to another. On the other hand, subjects in the unrestricted condition can get wrong sets or inadequate methods, such as working with too few of the numbers and letters, and there is no guidance, from external sources, away from such self-defeating response tendencies. It had originally been thought that a degree of control over the subject's choices, as exemplified by the various restricted conditions, would be advantageous since the subject would be prevented, to some extent, from persisting too long on wrong hypotheses. The results merely suggest that extreme restriction (as in Group RNL where the subject had no choices to make), though inferior to the unrestricted condition, may have had some advantage over partial restriction (Groups RL and RN).

In the unrestricted condition, the percentages of correct and incorrect pairs may deviate considerably from the 50:50 split imposed upon the restricted conditions. As will be shown in Experiment II below, the number of correct pairs increases as subjects begin to discover the principle. It is known that subjects tend to prefer positive to negative instances (Hovland and Weiss, 1953), and to exhibit enumerative instead of eliminative behaviour (Wason, 1960). Thus, the restricted conditions force the subject to deal with more incorrect instances than he would probably like, and prevent him from achieving the "success" of a long string of positive instances.

## EXPERIMENT II

Examination of the pairs listed by Group U in Experiment I showed that a portion of the subjects apparently used some systematic approach in their selections of numbers and letters (such as pairing every letter with one number before changing to another number), while other subjects appeared to be quite unsystematic. Also, the fact that the numbers and letters were printed on the subject's sheet in normal order suggested, to certain subjects, various systems to be tried out. Experiment II is in part an attempt to learn something more about the role of systematic vs. unsystematic approaches in inducing a principle. In addition, certain analyses of the instances will be reported for what they may reveal about the induction process.

### METHOD

The task, instruction, and procedure were the same as for Group U in Experiment I, with the following exceptions. There were three forms of the sheet given to the subject, on which he wrote number-letter pairs. One form was the same as in Experiment I; the numbers and letters were listed in horizontal rows at the top of the sheet in normal order. For the other two forms of the sheet, both numbers and letters were also listed in rows, but in random orders, with a different random order for each sheet. The orders were: 3 8 6 5 2 7 1 4 with D F C A E H B J G I, and 6 1 8 3 7 5 4 2 with H E G D J C I A F B. The three sheets were handed to the subjects in turn so one-third of the subjects was run with each sheet.

Any subject who had not discovered the principle in 20 min. (vs. 25 min. in Experiment I) was recorded as a nonsolver. At the end of the experiment, but while the subject still had his sheet in front of him, the experimenter questioned every subject, solver or nonsolver, as follows: The experimenter said, "Now would you go back to the first few pairs you wrote down (the experimenter pointed toward the first two or three pairs). I would like to ask you why you selected the pairs you did when you first started. Your first pair was (number and letter were stated). Try to remember what you were thinking about. What I'm trying to find out is, did you have any plan, any reason, no matter how vague, for choosing the first few pairs you did." The experimenter took notes on what the subject said.

The subjects were 135 students from introductory psychology classes.

## RESULTS

There were 75 subjects who discovered the principle. The remaining 60 subjects were nonsolvers. The analyses to be reported chiefly involve comparisons of these two groups, solvers vs. nonsolvers. It may be noted in passing that the task was easier when the numbers and letters were listed in normal order; 35 of the 45 subjects with the normal-order sheet solved, whereas 40 of the 90 subjects with random-order sheets solved.

In the postexperimental inquiry, 43 solvers and 21 nonsolvers reported that they had had some plan or system in mind when they began to write down pairs. Of those subjects who said they had had no system, 32 were solvers, 39 were nonsolvers. Thus, although the differences are not great, the two largest frequencies are systematic solvers and unsystematic nonsolvers. In a test of independence of the four categories, $\chi^2 = 5{\cdot}81$, $p < 0{\cdot}05$, 1 *d.f.*

Incorrect attempts to state the principle were recorded for both solvers and nonsolvers. The mean for solvers was 1·13; for nonsolvers, the mean was 1·28. The difference is not significant. Furthermore, the frequencies are so low (by far the most frequent number was 1) that no relationships could be found between incorrect attempts and the measures reported below.

Differences between solvers and nonsolvers show up in certain measures based on the pairs (of numbers and letters) the subjects wrote. Even though each solver worked only until he discovered the principle, while all nonsolvers kept working for 20 min., solvers wrote a mean of 38·0 pairs, while for nonsolvers the mean was only 31·6 pairs. The difference between the means yields $t = 1{\cdot}89$. If one nonsolver with a score of 137 pairs is excluded, the nonsolvers' mean becomes 29·8, and the $t$ becomes 2·81.

The proportion of the pairs that were correct, were instances of the principle, was computed for each subject. The mean proportion for the solvers was 0·62, for nonsolvers the mean was 0·56. These means differ significantly; $t = 3{\cdot}00$.

A final measure expresses subjects' tendency to vary only one factor at a time. As the subject writes down a pair, he can use both a different number and a different letter from the immediately preceding pair ("double change"), or he can vary only one of these elements by continuing to use the same number or letter he used in the preceding pair. For each subject the proportion of changes over successive pairs that were double changes was determined. The mean proportion was 0·45 for solvers, 0·61 for nonsolvers. The difference is highly significant; $t = 5{\cdot}03$.

## DISCUSSION

The differences found, in Experiment II, between those subjects who did, and those who did not, discover the principle are, of course, only associated with solving or not solving; causal status is not implied. With this limitation, the results can be examined for what information they provide about the task of inducing a principle from positive and negative instances.

It is clear that an approach that has some plan or system (if we can take the subject's word as to whether or not he had a plan) is not a great deal more successful than the reported absence of a systematic approach. The difference between solvers and nonsolvers on this characteristic was not great. One problem is that some plans actually make it more difficult for the subject to discover the principle, especially if the subject persists on a wrong hypothesis.

The fact that solvers wrote somewhat more pairs in less time than did nonsolvers shows, of course, that solvers' rate of overt response was faster. Perhaps nonsolvers spent too much time thinking instead of writing down pairs (although they did not make fewer attempts to state the principle). It has recently been argued (Duncan, 1963) that taking time to stop and think is worthwhile for some aspects of problem solving, but not for all.

A higher proportion of the number-letter pairs written by solvers were positive instances of the principle than was the case for nonsolvers. But this difference might have occurred simply because solvers find out, at some point, which pairs are positive instances. However, examination of the pairs at various points in practice showed that solvers had a slightly higher proportion of positive instances than did nonsolvers throughout the entire session. In their last few pairs, just before they stopped and announced the principle, solvers showed a sharp increase in proportion of positive instances (mean of 0·79 in the last tenth of instances), thus apparently exhibiting enumerative behaviour as described by Wason (1960).

Perhaps the most revealing difference between solvers and nonsolvers was in the "double change" measure. There was a clear difference in favour of solvers in tendency to hold one element, number or letter, constant in successive pairs, and vary only the other element. Examination of the pairs at various stages revealed that early in practice, both solvers and nonsolvers were showing mean proportions of double changes of approximately 0·60–0·65. As their over-all mean (0·61) shows, nonsolvers remained at about this level throughout the session. But in solvers, the proportion dropped rapidly from this high until in the last half of the session solvers were averaging only about 0·36 double changes. Thus, in their attempts to induce a simple law, solvers learned to use, and made much more use of, the advantageous device of varying only one thing at a time.

A final point seems worth noting. The present task, dealing with numbers and letters, permits subjects to entertain many wrong hypotheses. Despite the fact that one aspect of the subject's behaviour is overt (the pairs he writes down), it is clear, from the subject's incorrect attempts to state the principle, that there is very little control over his thinking. He is free to consider, implicitly, an almost infinite variety of ways that numbers and letters might be paired. This large population of wrong hypotheses seems to be the chief reason many subjects do not discover the principle.

This study was supported in part by Public Health Service Research Grant M-3182 from the National Institute of Mental Health. Thanks are due to Mr. I. David Isaacs for assistance in the study.

### References

Duncan, C. P. (1963). Effect of instructions and of information on problem solving. *J. exp. Psychol.*, **65**, 321–7.

Hovland, C. I., and Weiss, W. (1953). Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, **45**, 175–82.

Wason, P. C. (1960). On the failure to eliminate hypothesis in a conceptual task. *Quart. J. exp. Psychol.*, **12**, 129–40.

Wetherick, N. E. (1962). Eliminative and enumerative behaviour in a conceptual task. *Quart. J. exp. Psychol.*, **14**, 246–9.

*Manuscript received 28th June, 1963.*

# A HULLIAN REASONING SITUATION

BY

W. S. ANTHONY

*From the Psychology Department, University of Auckland, New Zealand*

An experiment on a well-known but little-investigated situation proposed by Hull, is reported. The results are inconclusive.

## INTRODUCTION

The situation briefly indicated in Figure 1 was proposed by Hull (1935, p. 221 *seq.*) to illustrate his $r_g$ theory of reasoning. The situation has been discussed by Hilgard and Marquis (1940), Osgood (1953), Deutsch (1956*a*, 1956*b*, 1960), Thomson (1959), Treisman (1960), and Broadbent (1961). In this literature, frequent reference is made to early experiments by Maier (1929); Maier's experiments inspired the situation proposed by Hull, but differed from the situation in many ways (Anthony, 1958).

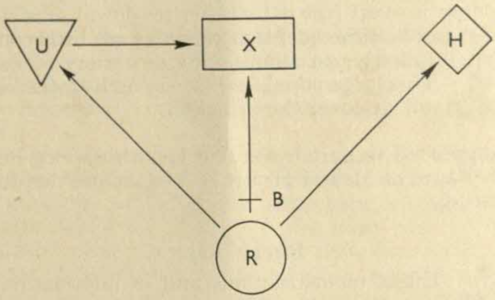


FIGURE 1

There is food at X, and water at U and H. The subject (rat) is first given training on the separate paths. On the test trial, he is placed at R, hungry, with the direct path to food blocked at B. The rat now has to choose between the detour-path RUX and the blind alley RH.

Three experiments by Wolfe and Spragg (1934), however, follow the situation more closely, and were known to Hull, but have been overlooked by others. Wolfe and Spragg obtained negative results: i.e. on the first test trial, five subjects chose correctly (RU) and six incorrectly (RH), on the second test trial, seven correctly and five incorrectly, and on the third, six correctly and six incorrectly. Eleven rats were used in the first experiment, and the same subjects, plus one more, were used in the second and third. Two deviations from the Hullian procedure were, in the order of training, and in the absence of reward at U or H in training.

The present experiment, already reported in an unpublished thesis (Anthony, 1958), is closer to the Hullian situation than any other done by, or known to, the writer. One group of rats is subjected to the situation; another group is allowed to explore the whole maze in training; "reasoning" or "assembly of behaviour segments" (Hull's term) is thus demanded of the former but not the latter group.

## METHOD

### *Apparatus*

The apparatus is shown in Figure 2. It does not follow Hull's suggestions in detail. The alleys were 4 in. wide, with walls 5 in. high, constructed of steel components, and floors of brown paper covering the table on which the apparatus was set. The walls of the four areas R, X, A and B, were 5 in. high, of aluminium or steel, covered at R, X and A, with other materials. Cream-coloured sandpaper covered the wall and floor at R.

Dark green baize covered the wall at X, and also the floor on top of a layer of cotton wool. Grey paper covered the walls and floor at A. At B, the walls were bare aluminium bent into vertical folds, and smooth aluminium covered the floor. A and B were designed to induce distinct overt responses: at A, a fibreboard barrier, 1 in. high, ½ in. wide, induced a getting-over response; the barrier was fixed at each end to two fibreboard ledges, 1 in. in height and width, which ran along each wall at A. At B, an overhead vertical aluminium sheet, leaving a gap 2½ in. high, induced a creeping-under response.

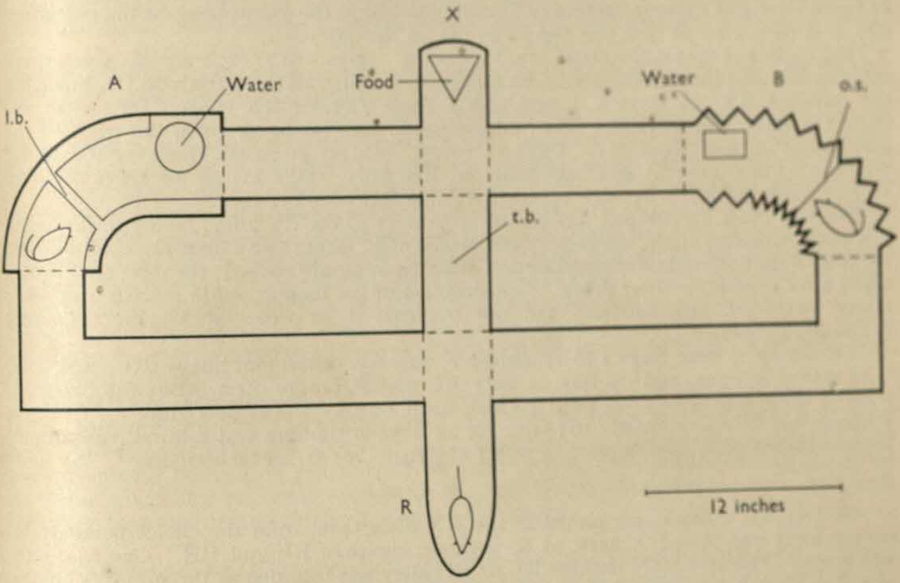


FIGURE 2

The dotted lines define the boundaries between each "box" (R, X, A, and B) and its adjacent alleys, and so also define the positions of blocks when in use. For any subject, either BX or AX is permanently blocked; thus for training on "UX," the subject is placed either at A or at B, as drawn. For all other trials, the subject is placed at R, facing as drawn. On the test trial, there is a block at *t.b.*, and the correct path is the one leading to the box which functioned as "U" for that subject. *l.b.*: low barrier. *o.s.*: overhead sheet. The *l.b.* and *o.s.* induce special responses in the passing subject.

At A, a white circular pottery dish contained water; at X, a wooden triangular dish contained food pellets; at B, a glass rectangular dish, set in an aluminium base, contained water. Removable walls, similar in appearance to the fixed walls of A, B and X, permanently blocked off BX (to make A function as "U" and B as "H") for one half of the subjects, and permanently blocked off AX (to make B function as "U" and A as "H") for the other subjects. In accordance with the procedure, alleys were also temporarily closed with hardboard blocks extending the full width and height of the alleys.

The maze was covered with perforated zinc sheeting, and was centrally placed in a room with bare white walls; lighting was diffuse, by reflection from the walls; the experimenter stood centrally near X; external cues would not be excluded.

### *Subjects*

Twenty-four experimentally naive hooded rats, 3 months old at the start of the experiment. Eighteen were male; six were female.

### *Training Procedure*

The subjects of each sex were randomly divided to make up two groups, each group containing nine males and three females. Group 1 was treated in accordance with the Hullian reasoning situation. Group 2 was treated in the same way as Group 1 except that exploration of the whole maze (but not of "HX") was allowed on each trial. For six members in each group, A served as U and B as H, and for the other six, B served as U and A as H.

The subjects were given six training trials per day for 22 days. They were trained under hunger on Days 1 to 10 inclusive and Day 21, and under thirst on Days 11 to 20 inclusive and Day 22. The test trial was on Day 23, under hunger. For hunger, the subject was given 1½ hr. of standard food immediately after its use in the experiment on the preceding day, with water constantly available. For thirst, the subject was given 1½ hr. of food and water immediately following its use in the experiment on the preceding day, and then 2 hr. of food just before its use on the day itself.

The nature of the training trials was as follows. Days 1 to 7: each rat was given three trials on UX and three trials on RX; on Days 1, 3, 5, and 7, all three trials on UX preceded the three on RX; on Days 2, 4, and 6, UX trials followed RX trials. Days 8 to 10: three UX trials and three RX trials, intermingled, each rat being given the trials in the same order on any given day. Days 11 to 15: each rat given six trials on RU daily. Days 16 to 20: each rat given six trials on RH daily. Day 21: as for Days 8 to 10. Day 22: each rat given a choice between RU and RH on six trials.

For UX trials, the subject was placed in U, next to the RU alley, facing towards the UX alley; for other trials, the subject was placed at R, facing away from the choice-point. Subjects were removed after consumption of the appropriate reward; the next trial would begin after a variable short delay. Consumption of the inappropriate reward was, very rarely, observed, although this had not occurred in previous work with the same deprivation schedule.

For Group 1, from Days 1 to 21 inclusive, only one of the four alleys (RU, RX, RH, UX) was ever open, and on Day 22 only RU and RH were open. But for Group 2, none of these four alleys was ever blocked, until Day 23. (The procedure for Group 2 subjects was therefore exactly the same for all days of training under thirst, the subject being removed after consumption of water at either U or H, on each trial.)

### *Test trial*

On Day 23, a block was placed in the RX alley, 4 in. from the choice-point at R; each subject was placed, hungry, at R, with the choice of RU and RH. One trial only was given. The first entry into the RU or RH alley was recorded as the subject's choice, although all subjects eventually reached X by way of R–U–X, and were removed after consumption of food there.

## RESULTS

On the test trial, of the subjects who chose RU, all went on to X and ate food there, except one subject in Group 2, who returned from U to R and explored further, finally reaching X and eating there. All the subjects who chose RH reached H, returned to R, entered RU, went on to X, and ate there. No subject drank water.

The results are analysed in two ways: first, the frequencies of RU and RH choices on the test trial, are compared; secondly, the test trial choices are compared with the control choices under thirst (Day 22).

On the test trial, in Group 1, nine subjects chose correctly (RU), and three chose incorrectly (RH). In Group 2, five chose correctly, seven incorrectly. In neither group is there a significant preference for RU or RH; for Group 1, $p = 0{\cdot}146$, two-tail "binomial test" (Siegel, 1956).

On Day 22, of Group 1 subjects, six chose RU more often than RH, three chose RH more often, and three chose RU and RH equally often. Six Group 1 subjects' choice on the test trial did not follow their preference on Day 22 (this includes the three who had shown no preference); four of these subjects changed in the correct direction (choosing RU), two in the incorrect direction (choosing RH). Applying the (binomial) "sign test" (Siegel, 1956), the total change is clearly insignificant. A similar analysis for Group 2 is omitted because these subjects sometimes ran RXU to water on Day 22, ignoring RU and RH.

Comparing choices of RA and RB (alleys leading to A, B) suggests an over-all preference for RA. On the test trial, seven Group 1 subjects chose RA, five, RB; in Group 2, eleven chose RA, one, RB. On Day 22, in Group 1, seven preferred RA, two, RB; in Group 2, ignoring RXU runs, seven preferred RA, three, RB.

In training, Group 2 subjects could run RUX. But the median total number of RUX runs per subject was only six (range, one to ten), all these runs being on days of hunger. On Day 21 (last day of hunger before test trial), seven out of 12 Group 2 subjects ran RUX at least once. Of these seven, only three chose correctly on the test trial.

## DISCUSSION

It is usually proposed that the criterion of "reasoning" in the situation is, that the subjects should choose RU more often than RH, on the test trial. This criterion is used by Wolfe and Spragg (1934), other writers, and Hull (1935, p. 226); but another criterion stated by Hull (1935, p. 222–3) is that RU must be chosen more often on the test trial (subjects hungry) than on the control trials (same subjects, thirsty). The second criterion draws attention to the expectation that the rat will discriminate between hunger and thirst. Our result for Group 1, the relevant group, is negative by both criteria, although in the right direction; the result for Group 2 is also negative by the first criterion, the second being inapplicable to Group 2.

Maier (1935) suggested that the negative results of Wolfe and Spragg might be due to the block in RX on the test trial; however, the rat can apparently succeed in the Tolman-Honzik insight situation (evidence reviewed by Anthony, 1959), where a block is imposed in the test trial. Another possible difficulty in the Hullian situation arises from the use of blocks in training (this would not apply in our Group 2). Up to the test trial, paths leading to the inappropriate reward have always been closed (Hull, 1935, p. 222). This means that (*a*) the subject has never run down a path and *not* found the appropriate reward; (*b*) whenever RU has been open, UX has been closed, so that the subject could logically doubt whether the whole path R–U–X would be open on the test trial. Hypotheses about the disturbing effects of blocks on drive discrimination learning, were confirmed by two experiments (one reported by Deutsch and Anthony, 1958, and both by Anthony, 1958).

The Hullian reasoning situation raises theoretical problems similar to those raised by latent learning situations. Positive results on various latent learning situations have been obtained and generally accepted. In contrast, the Hullian situation appears to be too difficult for the rat. The failure of both our groups, those set the Hullian situation *and* those given free exploration, suggests that the Hullian situation presents even more difficulty than just the assembly of behaviour segments.

A systematic error in the analysis of results in the unpublished report (Anthony, 1958) of the present experiment, has been corrected here. The experiment was done at the Institute of Experimental Psychology, Oxford University. Acknowledgement is due to Dr. J. A. Deutsch for supervision, and to the Medical Research Council for a research training grant.

*Additional note:* After this article was submitted, Dyal and Jones (1963) and Dyal and Abright (1963) reported experiments on the situation, and referred to other published work. All experiments have yielded negative results on the test trial, in accordance with the present writer's results. Two distinctive features of the present study are: (1) apparatus for inducing distinctive overt responses at U and H; (2) the inclusion of a control condition in which free exploration was allowed.

## REFERENCES

ANTHONY, W. S. (1958). *The Assembly of Behaviour Segments.* Unpublished D.Phil. thesis, Univ. Oxford.
ANTHONY, W. S. (1959). The Tolman and Honzik insight situation. *Brit. J. Psychol.*, **50**, 117–24.
BROADBENT, D. E. (1961). *Behaviour.* London.
DEUTSCH, J. A. (1956*a*). A theory of insight, reasoning, and latent learning. *Brit. J. Psychol.*, **47**, 115–25.
DEUTSCH, J. A. (1956*b*). The inadequacy of the Hullian derivations of reasoning and latent learning. *Psychol. Rev.*, **63**, 389–99.
DEUTSCH, J. A. (1960). *The Structural Basis of Behaviour.* London.
DEUTSCH, J. A., and ANTHONY, W. S. (1958). Blocking the incorrect alley in a two-drive learning situation. *Quart. J. exp. Psychol.*, **10**, 22–8.
DYAL, J. A., and ABRIGHT, J. (1963). Problem-solving assembly of separately acquired behaviour segments: negative results. *Psychol. Rep.*, **13**, 619–21.
DYAL, J. A., and JONES, C. A. (1963). Experimental tests of Hull's reasoning paradigms. *Psychol. Rep.*, **13**, 583–90.

HILGARD, E. R., and MARQUIS, D. G. (1940). *Conditioning and Learning*. New York.

HULL, C. L. (1935). The mechanism of the assembly of behaviour segments in novel combinations suitable for problem solution. *Psychol. Rev.*, **42**, 219–45.

MAIER, N. R. F. (1929). Reasoning in white rats. *Comp. Psychol. Monogr.*, **6**, No. 3.

MAIER, N. R. F. (1935). In defence of reasoning in rats: a reply. *J. comp. Psychol.*, **19**, 197–206.

OSGOOD, C. E. (1953). *Method and Theory in Experimental Psychology*. New York.

SIEGEL, S. (1956). *Nonparametric Statistics for the Behavioural Sciences*. New York.

THOMSON, R. (1959). *The Psychology of Thinking*. London.

TREISMAN, M. (1960). Stimulus-response theory and expectancy. *Brit. J. Psychol.*, **51**, 49–60.

WOLFE, J. B., and SPRAGG, S. D. S. (1934). Some experimental tests of "reasoning" in white rats. *J. comp. Psychol.*, **18**, 455–69.

*Manuscript received 23rd July, 1963.*

# CONCERNING THE PSYCHOPHYSICAL POWER LAW

BY

S. S. STEVENS

*From the Laboratory of Psychophysics, Harvard University*

It is fortunate that, despite our theorizings, most of our facts manage to stay stubbornly as they were. I am encouraged by that thought as it bears on a recent theoretical attempt by Treisman (1964) to do over the psychophysical law. I should like, more in explication than in rebuttal, to comment on three major points relating to the law in question.

The first point concerns the body of interconnected observations that comprise the psychophysical law. The unaltered facts of interest here relate to observed correlations between inputs and outputs under a wide variety of procedures. The inputs are usually sets of graded stimuli. The outputs are responses made by human observers who, in accordance with instructions, may do one of several things, of which the following are examples. (1) *Cross-modality matching:* the observer adjusts a stimulus in one sense modality to make it appear as equal as possible to a given stimulus in another modality. (2) *Ratio production:* the observer adjusts a stimulus to make it appear to stand in a prescribed ratio to a given stimulus. (3) *Ratio matching:* the observer adjusts one of a pair of stimuli to make the ratio between the two match the apparent ratio between two other stimuli. (4) *Magnitude production:* the observer adjusts a series of stimuli to match a series of numbers. Most of these procedures have, as useful complements, procedures in which only the experimenter adjusts the stimuli. A combination procedure, recently tried out, allows the observer to perform both operations: he is asked to adjust the stimulus to a wide range of levels of his own choosing and to estimate the apparent magnitude of each stimulus that he produces.

Under all these procedures, in varied and repeated experiments, the correlation between stimulus and response has turned out to be a power function. The exponent of the power function varies from sense modality to sense modality. It also varies with such factors as state of adaptation.

People commonly use the word sensation in this context, and most observers feel certain that they are gauging the strength or intensity of their sensations when, for example, they report the relative magnitude of two loudnesses or of two brightnesses under instructions to do so. Usage based on the foregoing operations leads us to make shorthand statements, and to say that sensation intensity grows as stimulus intensity raised to a power. This way of speaking must not obscure the fact that the psychophysical power law rests on the operations set forth above. The terms related are observables, and these observables are correlated by power functions.

As it turned out, the wide invariance of this relation (Stevens, 1961, 1962) astonished me, as it no doubt astonished others, for the existence of a simple law relating stimulus magnitude to judged magnitude on more than two dozen perceptual continua was hardly to be expected. With the finding (Stevens and Guirao, 1964) that apparent viscosity grows as the 0·5 power of the physical viscosity, we may be nearing the end of the inventory of continua on which the power law can be tested. Though we may be running out of easily testable continua, no firmly established exception has yet appeared to limit the generality of the psychophysical power law, which is an expression of the principle that equal stimulus ratios produce equal sensation ratios.

These stubborn facts, the correlations between observables, provide the foundation for and give basic meaning to the psychophysical power law. Treisman proposes a different definition of the psychophysical law. He would have the law refer, not to the correlation between two observables, but to the relation between one observable (the stimulus) and a hypothetical entity labelled S. The entity S is a "neural effect" that resides somewhere between the observable stimulus and the observable response. Treisman then shows how he can alter his particular variety of psychophysical law at will, by merely making one or another assumption about how the entity S is related to response. Thus, given that response is a power function of stimulus, if one wants the entity S to be a logarithmic function of the stimulus (and thereby rescue Fechner?), all one need do is decide that response shall be an exponential function of S.

No doubt many different transfer functions make themselves felt along the chain between input and output in a complex organism. Until we know the functions, we are free to try to guess their characteristics. But one wonders whether Fechner would want to be proffered rescue by a thread so frail as a conjecture about the exponential dependence of responses on neural effects. Perhaps we do better to stick to observables.

The second point concerns an error made by Treisman, as well as by others. By failing to catch the error he is led to imply that cross-modality matching has not succeeded in its goal of closing a vital circle in a process of validation.

When stimuli in two sense modalities are matched in apparent magnitude, the matching function is demonstrably a power function. This outcome is consistent with both of two possibilities: the two modalities could be governed either by power functions or by logarithmic functions. This latter possibility lures Treisman into supposing that cross-modality matching justifies the logarithmic function as well as the power function. But there remains the question of the exponent of the matching function. Here we find a crucial difference. If we know the power functions governing the two sense modalities, we can predict the exponent of the matching function, but prediction is not possible with logarithmic functions, for the exponent of the matching function then depends only on arbitrary constants. For a good discussion of this issue, see Luce (1963).

It is important to be clear about the cross-modality experiments, because they appear to constitute the surest road to an objective psychophysical law. Since cross-modality matching now appears to be the most important of the methods, I can only wish I had hit upon it much earlier. With that method we could have learned all that we need to know about the over-all operating characteristics of the sensory systems without ever asking an observer to make a numerical assessment of any kind. Since subjective numerical estimations are still suspect to many scientists, reliance on cross-modality matching could perhaps have steered us clear of some kinds of misunderstanding.

Of this, at least, we can now be reasonably sure: the *relative* values of the exponents of all the perceptual continua can be obtained on demand without the observer's ever having to use numbers to report his judgment. If *absolute* values of the exponents are desired, then, as in all measurement, we will need to agree on a reference or standard. What continuum should we choose? If we were to forego the use of numbers for the reference continuum, I should nominate apparent length of lines as a plausible substitute. With the aid of a device for projecting variable lines of light, we have successfully scaled both brightness and loudness in terms of length of lines (Stevens and Guirao, 1963). On the assumption that the exponent for length is unity, the exponents obtained by adjusting lines to match brightness and loudness agree with those obtained with numerical estimation.

The main drawback to limiting the arsenal of methods to a single procedure is loss of convenience. People carry numbers around in their heads, so to speak, and they can emit them on demand. It is less easy to resort to a darkened room, a screen and a precision line projector every time a continuum needs to be scaled. In actual practice we usually employ both procedures, cross-modality matching plus the various numerical procedures. If you care about an exponent, you can probably pin it down tighter if you use a variety of procedures. It helps to balance out the biases.

The third and final point concerns the problem of "explaining" the psychophysical power law. Here we enter the realm of opinion, but interest in this problem has led to a variety of proposals. Perhaps no one of them yet constitutes a definitive account, but some hypotheses have evoked much interest and discussion. Notable in particular is MacKay's suggestion (1963) of a "comparator" mechanism by which the impulses from a sense organ are balanced against the output of a central generator of some kind. This conception has much novelty, for it casts the central nervous system in the role of active participant in a process of weighing or counterbalancing the sensory input. Moreover, the power-law response of the over-all system would eventuate for a wide variety of transfer functions in the elements of the system. Whether the nervous system works according to this plan is, of course, not known.

Equally moot is the explanation based on learning. Treisman seems to favour the view, suggested by many, that the power functions are learned by the organism as it adapts to its world. Those who choose to oppose the idea that sensory systems have built-in operating characteristics apparently believe that the organism starts essentially blank, and that the child learns through adaptive adjustments to gauge the stimulus intensity and to make judgments consistent with a power function having the appropriate exponent. In order to make this hypothesis sound plausible, ingenuity has been supplied in varying measure. To an older generation it was proposed that people learned to judge a doubled apparent brightness because they had seen rooms lighted by one candle and by two. On this basis the brightness exponent would be 1·0. Recently a more roundabout road has been proposed. People are said to learn by noting how brightness varies with the distance of a reflecting surface from a point source. If they have also learned that distance has the exponent 1·0, then the operation of the inverse-square law would allow them to learn that brightness has the exponent 0·5. Several laboratories have found, however, that the exponent for brightness lies in the vicinity of 0·33.

Treisman goes so far as to suggest that the operation of the inverse-square law "explains" the exponent for olfaction, which for many substances lies in the vicinity of 0·5 or 0·6. How, one wonders, do we learn to use the inverse-square law in sniffing odours when they are so often borne on air currents?

This indeed is the point. Learning may well account for the values of some of the power-function exponents; on present evidence we cannot say yes or no. It seems improbable, however, that the principle of learning can cover all sensory continua. Such a conjecture requires that one go out of his way to deny the existence of biological differences in sensory systems. For most continua it seems simpler to accept as a working hypothesis the view that the transfer functions differ from one sense modality to another.

The power function applies to the over-all system, a system of many stages and many transductions. When the exponent of the power function differs from unity, we do not necessarily know where in the system the nonlinearity was introduced, but it is not improbable that the peripheral mechanisms are sometimes implicated. For one thing, we can show that, by changing the state of adaptation of a periphral transducer, the power function can be altered (J. C. Stevens and S. S. Stevens, 1963). When we light-adapt the eye to a high level, the exponent increases from about 0·33 to about 0·44. Other parameters of the power function also change when the state of the sense organ is altered.

Many points other than those noted above are touched upon by Treisman as he conjectures this and that. The psychophysical power law has been documented further since Treisman wrote, and there is no need to debate matters that have been or should be settled in the laboratory. Our important debates are not with one another; they are with nature.

## References

MacKay, D. M. (1963). Psychophysics of perceived intensity: a theoretical basis for Fechner's and Stevens' laws. *Science*, **139**, 1213–6.

Stevens, J. C., and Stevens, S. S. (1963). Brightness function: effects of adaptation. *J. opt. Soc. Amer.*, **53**, 375–85.

Stevens, S. S. (1961). The psychophysics of sensory function. In Rosenblith, W. A. (ed.). *Sensory Communication*. New York, pp. 1–33.

Stevens, S. S. (1962). The surprising simplicity of sensory metrics. *Amer. Psychologist*, **17**, 29–39.

Stevens, S. S., and Guirao, M. (1964). The scaling of apparent viscosity. *Science*, **144**, 1157–8.

Treisman, M. (1964). Sensory scaling and the psychophysical law. *Quart. J. exp. Psychol.*, **16**, 11–22.

*Manuscript received 28th April, 1964.*

# A NOTE ON TREISMAN'S MODEL

BY

J. P. N. PHILLIPS

*From the Department of Psychology, The University, Hull*

It is argued (i) that Treisman's model specifies not two but an infinite number of psychophysical functions, (ii) that the additional assumption made to decide between psychophysical laws is completely arbitrary and (iii) there is no contradiction between Fechnerian and Stevensonian psychophysics.

In an attempt to reconcile the "intractable and inconclusive dispute" between Fechnerian and Stevensonian psychophysics, Treisman (1963, 1964) has put forward a model according to which, in a scaling experiment, a stimulus I determines a neural effect $S = f(I)$, from which the subject's response $R = g(S)$ is computed. He argues that Stevens' finding that $R = kI^n$ can be explained, by appropriate choice of $g(S)$, either on the hypothesis that $S = k' \log I$ or that $S = k''I^n$. However, there are not two, but infinitely many possible hypotheses. For let f be any function with an inverse: then setting $R = g(S) = k[f^{-1}(S)]^n = kI^n$ will account for Stevens' results. Thus, unless $g(S)$ is specified, there will be an infinite number of possible functions f. (Whilst the results of cross-modality matching experiments exclude an infinite number, they still leave an infinite number of possible f's.)

To derive different predictions from the logarithmic and power functions, Treisman makes the further assumption that more errors will occur in experiments where the calculation of $g(S)$ requires "complex" computations (e.g. extraction of roots) than where it requires "simple" ones (e.g. addition). However, this is completely arbitrary, and the reverse assumption might equally well be made: whilst for many desk calculators the extraction of roots is more complex than addition, precisely the reverse is true of many slide rules. In the absence of evidence about the type of the posited computer, no differential predictions can legitimately be made in this way.

Thus Treisman's model allows an infinite number of psychophysical functions, and does not make possible a decision between them, in particular between the logarithmic and power functions. It may be doubted whether such a decision is necessary, i.e. whether there is any real contradiction between the two psychophysics. Fechner found that if numbers are assigned to stimuli in such a way that the difference between the numbers assigned to equally often discriminated stimuli is constant, then the numbers are a linear function of the logarithms of the stimulus intensities: Stevens, that the numbers uttered by suitably instructed subjects to stimuli are a power function of them. Although they indicate different conventions for assigning numbers to stimuli, these two propositions are mutually consistent.

## REFERENCES

TREISMAN, M. (1963). Laws of sensory magnitude. *Nature*, **198**, 914–915.

TREISMAN, M. (1964). Sensory scaling and the psychophysical law. *Quart. J. exp. Psychol.*, **16**, 11–22.

*Manuscript received* 18*th August*, 1964.

# WHAT DO SENSORY SCALES MEASURE?

BY

MICHEL TREISMAN

*From the Institute of Experimental Psychology, University of Oxford*

It is indeed an honour to have drawn Professor Stevens' fire (1964), in view of the magnitude of his research achievement in the field of sensory scaling. If I shall differ with him on certain of his conclusions, it is in full awareness of the debt we all owe him for the experimentation and thinking which has provided much of the factual basis and theoretical stimulus to further research and development in this area.

Professor Stevens (1960, 1964) advocates a psychophysical power law, while I have questioned it (Treisman, 1961, 1962, 1963*a*, 1963*b*, 1964*a*). I shall here examine the main points of disagreement lying behind these differing positions.

### (1) *What are the "observations"?*

Stevens (1964) states that it is "the correlation between stimulus and response," a relation between "observables," which gives the power law, and that "to say that sensation intensity grows as stimulus intensity raised to a power" is a "shorthand statement" of this. But this does not seem to me a fair description of what actually happens. Let us consider a "direct" scaling experiment in which the method of fractionation is applied to lifted weights (W). A standard weight ($W_s$) is presented to the subject, together with a range of comparison weights. The instructions ask him to choose that one of the comparison weights which feels half as heavy as $W_s$. He chooses such a weight, $W_c$, and the procedure is then repeated with a new standard weight. Eventually a number of corresponding pairs of weights ($W_c, W_s$) will have been selected.

Now the only "observables" that emerge from such an experiment are the paired weights ($W_c, W_s$). The experimenter has observed nothing else that he will use in deriving the psychophysical law. And no power law relates these observables. Indeed, it can be shown that if the psychophysical law is to be a power function, then the relation between $W_c$ and $W_s$ must be linear (Ekman, 1958), i.e. we will have $W_c = kW_s + b$, where b is a constant. (For simplicity of presentation, b and other such constants will be dropped when they do not affect the arguments presented.)

No power law is revealed by these observables. But such a law will emerge if we further assume (*a*) that there is a "psychological magnitude" or dimension or function of W which I shall refer to as "S," and which Stevens sometimes calls "$\psi$," such that $S = f(W)$, $S_s = f(W_s)$ and $S_c = f(W_c)$, and (*b*) that since the instructions ask the subject to halve something, and since he has not halved $W_s$, what he has halved is $S_s$. Then we shall have $\frac{S_c}{S_s} = \frac{1}{2}$, and $\frac{W_c}{W_s} = k$, and we can write $\left(\frac{W_c}{W_s}\right)^n = k^n$, where n is a constant such that $k^n = \frac{1}{2}$ (i.e. $n = \log\frac{1}{2}/\log k$). It follows that we can write $S = cW^n$, where c is a constant determining the unit of S. Here we have the power law, and the first major point of disagreement between Stevens and myself is that I would say the experimenter has assumed S, and has inferred its relation to W, but Stevens considers that it is an "observable."

How could such a claim be justified? Perhaps by the supposition that as well as observing W the subject also and at the same time "observes" something internal to himself which corresponds to S. Possibly this is meant to be the implication of Stevens' (1964) assertion that "most observers feel certain that they are gauging the strength or intensity of their sensations." But this is at best an analogical use of the word "observe." To identify the experimenter's observation of his data with an assumed relation between the subject's statements and something internal to him provides a shaky basis for a scientific argument (Treisman, 1962). In general we are reluctant, and wisely so, to accept what subjects say as factually unquestionable information about the causation of their behaviour. For instance, if a subject were told to watch a display and were then asked, "Was it continuously visible to you as you watched it?", he would probably answer "Yes," yet there is considerable evidence that the visual threshold is high during saccades and low during fixation pauses (Volkmann, 1962), and that the visual input is sampled in snatches about 100 millisec. long (White, 1963). Neither of these findings would find any support in the subject's introspections. There seems no reason why a subject's belief that he is, for example, halving a psychological magnitude, should not be treated with similar circumspection.

If S is not directly observed, then we must relate it both to the input, $W_s$, and to the output, in this case the weight which the subject selects as being half as great as $W_s$, i.e. $W_c$. This gives the framework of an explanatory model of the usual scientific sort (Braithwaite, 1953) in which some symbols, which are co-ordinated with the experimenter's observations ($W_s$,$W_c$) are related by defined rules to others (S) which intervene. As we have developed this model above we have the relations: $S_s = W_s^n$ (this corresponds to the so-called "psychophysical law"); $S_c = \frac{1}{2}S_s$ (the result of the operation performed on S to select a weight half as heavy as $W_s$); and $S_c = W_c^n$. This model includes a power function psychophysical law; could such a model be based on any other law?

(One feature of a psychological model is that intervening terms which do not refer directly to behavioural observations may nevertheless be capable of being co-ordinated to potential physiological observations. We have little doubt that the sensory effects of lifting a weight are conveyed by nerve impulses which travel along nerve fibres to the brain and that the response is mediated by similar impulses in motor nerves. It would seem oddly unparsimonious to suppose that the intervening stage, in which the sensory information is used to select the response, was not performed by further neural mechanisms in the brain, but in some other, mysterious, way. It was for this reason that I suggested that S may eventually be identified as a "neural effect" of the stimulus.)

### (2) *How many psychophysical laws?*

A scientific model is changed or discarded when its predictions or logical consequences disagree with observation. But if different models give the same predictions we cannot choose between them in this way. In this sense the model above is not unique. The same prediction ($W_c = kW_s$) will be given by a logarithmic psychophysical law if we describe the operation on the intervening scale as $S_c - S_s = \log \frac{1}{2}$. (This is no more or less arbitrary than $S_c = \frac{1}{2}S_s$, unless we suppose that our subjects have an inner knowledge into the proper forms our scientific theories should take.)

If $W_c = kW_s$, then $\log W_c - \log W_s = \log k$, and $n\log W_c - n \log W_s = n \log k = S_c - S_s$, where $n \log k = \log \frac{1}{2}$ (i.e. $n = \log \frac{1}{2}/\log k$). It follows that we can write $S = n \log W + \log c$, and we have n appearing as a slope constant but with the same value as when it serves as the exponent of the power law.

The point is that a model sufficient to account for the result of any direct scaling experiment can be based on either a power function or a log function law. This is true of each scaling procedure, not just of fractionation, when the model is adapted appropriately. In each case the information contained in an experimental relation, such as $W_c = kW_s$, must be spread over two theoretical functions, such as $S = f(W)$ and $S_c = g(S_s)$, and this information is sufficient to determine what pairs of functions are compatible, but not to specify any one pair uniquely. It is, of course, obvious that there are an infinite number of ways in which $S_c = g(S_s)$ could be specified, each of which would give a corresponding psychophysical law, as Phillips (1964) notes. The predominance of the logarithmic and power function laws in the literature probably reflects the simplicity of the corresponding relations between $S_c$ and $S_s$, which involve only either addition or multiplication.

It follows that no number of "experimental demonstrations" of the power law can show that the logarithmic law is wrong. Each such "demonstration" will have required an arbitrary, even if implicit, postulate such as $S_c = \frac{1}{2}S_s$. In each case the relation between $S_c$ and $S_s$ (or between R and S, depending on the procedure) could equally well have been so chosen that the same data would have "demonstrated" the logarithmic function. But if so, how is it that cross-modality matching has "validated" the power law?

### (3) *Does cross-modality matching "validate" the power law?*

In a cross-modality matching task the subject must pair stimuli in two modalities, such as brightness and loudness, so that they appear equally intense to him. Thus if lights of different brightness are presented, he will select a sound which seems to correspond in loudness to the brightness of each light, and it will be found that $I_1$ is a power function of $I_2$ (where $I_1$ is the intensity of a sound and $I_2$ the brightness of a light, in units such as dynes/cm.$^2$ and millilamberts) or, if the two intensities are measured in decibels, that they are linearly related.

This, of course, is the observable experimental result, and it is a matter of interpretation to suppose that what the subject "really" matches are quantities $S_1$ and $S_2$, related to $I_1$ and $I_2$ by power functions $S_1 = I_1^n$ and $S_2 = I_2^m$. If this assumption is made, we get $I_1^n = I_2^m$ and consequently $\log I_1 = \frac{m}{n} \log I_2$. This predicts the slope of the matching function, and if n and m have been determined by a direct method such as fractionation, the prediction is confirmed. But the power function is not the only route to this prediction: if fractionation results had been scaled on the assumption that $S_c - S_s = \log \frac{1}{2}$ then they would have given the log laws $S_1 = n \log I_1$ and $S_2 = m \log I_2$, with the slope constants having the same values as the exponents above; and the same prediction would, of course, have followed from these laws, and been confirmed. This is why cross-modality matching does not "validate" the power law as such.

The "error" that has not been "caught" in this case, both by Stevens and by Luce and Galanter (1963), whose argument is essentially the same, is the failure to appreciate that a log law derived from direct scaling data will have as its slope constant the same value that appears as the exponent of the power function given by those data. It is, of course, true that the slope of the matching function would not be predicted by log functions obtained by some other procedure, such as summing j.n.d's, which gives arbitrary slope constants, but it has not been proposed that it would (Treisman, 1964*a*). The scale derived from difference threshold data is not necessarily the same as that underlying magnitude judgement tasks; they may well represent different "neural effects" of the stimulus (Treisman, 1964*b*, 1964*c*).

(4) *Are psychophysical laws learned?*

"Measurement is the assignment of numerals to objects or events according to rules" (Stevens, 1951). This definition summarizes the modern appreciation that numbers do not inhere in objects or events; the ways we apply them require explanation. We are not born knowing how to use the words of French or English or any other language; we have to learn how to apply them to the situations and events we come across. And this must be as true of numbers as of other words. If different people employ the same rules when using numbers in a given situation it is most probably because they have been taught these same rules, whether explicitly, as when we learn to count, or in less obvious ways. It is difficult indeed to see how else we could explain our mutually comprehensible use of numbers and other evaluative terms.

On this hypothesis experience will have adjusted the "neural computer" which handles sensory information so that it will use the input from a given sense organ to select a number-word which is appropriate to the stimulus, in terms of the rules it has learned; and it will not particularly matter what function describes the characteristic of the "sensory transducer," as long as the sensory message has some reliable relation to the stimulus magnitude. It then seems more plausible that the subject in a scaling experiment is operating this same evaluative mechanism in slightly unusual circumstances, than that he is somehow short-circuiting the system in order to report directly the state of the neural machinery (impulse rate? concentration of a transmitter substance at a synapse? number of fibres firing?) at some point on the sensory input side. This hypothesis is supported by the relatively good agreement between the exponents found experimentally, and those predicted from the physics of the stimuli, for a number of modalities (Warren, 1958; Treisman, 1964*a*). This suggests that further attempts to relate exponents to the physical properties of the stimuli may prove useful, and that sensory scaling procedures may provide a technique for studying the processes by which we learn to use evaluative and quantitative words.

This argument does not imply that psychophysical methods cannot be used to make inferences about the physiological effects of stimuli—this can be done (Treisman, 1964*c*)—but that direct scaling procedures do not do this.

(5) *Which is the better psychophysical law?*

If the validity of the "introspective criterion" is rejected, it appears that a conventional sensory scaling experiment cannot establish a unique psychophysical law. This follows necessarily from the fact that determining the psychophysical law is equivalent to specifying a scale for S. But as "measurement is the assignment of numerals to objects or events according to rules," if we wish to scale S we must choose an appropriate convention or rule, and no experiment can tell us which of two possible rules is the "right" choice. Each will give a scale which can form the basis of a coherent descriptive system which can be applied to the results of any experiment. We cannot select between them by a crucial experiment, but this does not mean that we cannot make a choice. To do so we must apply the criteria by which we choose useful and readily manipulable systems of description, rather than clumsy and confusing ones.

A famous example was the conflict between the geocentric and heliocentric descriptions of the solar system. Each is equally valid, in the sense that any correct statement in one set of terms can be translated into a corresponding correct statement in the other. One is not "right" and the other "wrong"; nevertheless, the acceptance of Copernicus' system was a great advance. The Ptolemaic system was clumsy and confusing; the heliocentric description could encompass a great variety of observations in simple and coherent terms that led on to further insight. It was, of course, an empirical fact that the structure of the solar system led to this result; had it been differently constructed the geocentric description might have proved the simpler. But the choice would still have been determined by the utility of each description when applied to the empirical observations.

As a different example, consider that we are investigating an unknown device which can combine two input signals and give one corresponding output. We have adopted conventions which allow us to measure each signal on "$\text{Scale}_1$," and we find that when the two inputs measure 3 and 4 units on this scale the output is 5 units: "3 & 4 → 5." Similarly, "7 & 11 → 13·04," "5 & 12 → 13," etc. What light do these data throw on the device? One solution would be to accumulate a great many of these results, tabulate them, and use the tables to forecast the future performance of the device. Or we might observe that a different rule would define a "$\text{Scale}_2$" on which these observations would now be represented by: "$9 + 16 = 25$"; "$49 + 121 = 170$"; "$25 + 144 = 169$." Given this choice we should probably prefer the $\text{Scale}_2$ descriptive system, since it will allow us to identify the device as an "adder" and predict its future performance without more ado. (The relation between the scales is given by $\text{Scale}_1 = \text{Scale}_2^{\frac{1}{2}}$.)

These considerations suggest that to choose between the laws we should examine their application to the empirical data on scaling to see which descriptive system handles the data with the greater facility. I suggested that the human operator should be compared to a computer which measures the stimulus and then performs operations on this measure in accordance with a programme, and that we would prefer a description which makes the calculations it must do appear easy and simple when in fact it does perform readily and adequately, and difficult when its performance is slow and erroneous (Treisman, 1964*a*). Phillips (1964) has objected that this is "completely arbitrary" since what appears to us a complex operation, the extraction of roots, may be easier than a simple operation, addition, when a slide rule is used: in the absence of evidence about the type of computer we can make no predictions, he says.

The error in this argument is Phillips' assumption that a slide rule can be described in only one way. Let us consider a rule with two parallel scales, A and B, consisting of numbered corresponding points with the relation $A = \sqrt{B}$. We could renumber these points on scales $A'$ and $B'$ (given by $A' = \log A$, $B' = \log B$) and we would now have $A' = \frac{1}{2}B'$. Extracting the square root will have been replaced by the simpler operation of halving. In fact we do not do this because we should lose more than we gain: B is normally identical with an observation or measure of data, but $B'$ is related to it in a more complex way. We find it far simpler to leave a number unchanged than to find its logarithm; this seems so obvious that it never occurs to us to scale the slide rule and describe its use in terms of $A'$ and $B'$. In contrast, in the example above we preferred to call our device an adder rather than have the operations "&" and "→" because we were free to choose our scale. But if we had found that in many important applications $\text{Scale}_1$ gave simple descriptions but $\text{Scale}_2$ led to confusing complexity, then we might have resolved the conflict by retaining $\text{Scale}_1$ and defining the complex operations "&," the taking of a sum of squares, and "→," extraction of the square root.

Now it is conceivable that with the advance of physiological knowledge a similar conflict may appear between the description of the "neural computer" suggested by behavioural work, and the "natural" description of its neurophysiological constitution. For example, we would almost certainly prefer to count numbers of nerve impulses on the scale of positive integers, 1,2,3 . . . rather than, say, 0, 0·3, 0·48, 0·6. . . . But this might lead to an interpretation of $S = f(I)$ which gave a clumsy basis for describing behaviour. It would then be the case that in a single overall description some complexity was unavoidable, and we must choose where we would prefer to have it (unless, as with "wave" and "particle" descriptions in some areas of physics, neither system has a clear overall advantage in simplicity, and both are maintained in parallel). But a conflict of this sort has not yet been found and may never be found; its possibility does not make it "completely arbitrary" to prefer simple and helpful description at this stage.

It should also be made clear that the comparisons which led to the conclusion that the log law was likely to prove more useful did not depend especially on a preference for addition rather than the extraction of roots. For example, in discussing the prothetic-metathetic distinction the choice was between representing a situation where the environment provided no stimulus to change by either leaving a measure unchanged or imposing a logarithmic transformation on it. It was argued that on prothetic scales, such as loudness, the primary measure of the stimulus, S, is transformed by learning into a form, R, simply related to other stimulus information, such as the distance of the sound source. In the absence of similar correlated information metathetic continua, such as pitch, undergo no such change. Taking $S_f = \log I$ we can describe this by saying that in the absence of environmental pressure $S_f$ remains unchanged and is revealed by all the scaling procedures. Learning resulting from an environmental stimulus relation causes $S_f$ to be exponentially transformed. In this form the description is fairly plausible. But how will it look if we take $S_p = I^n$? We must now say that when there is no pressure from the environment this will cause $S_p$ to undergo a logarithmic transformation before magnitude estimates are made or category labels assigned. The effect of learning produced by correlations between stimulus inputs is that $S_p$ remains unchanged. It is obvious that if this account is correct, its description in $S_p$ form is less facile and more confusing than when $S_f$ is used as a basis for description.

## REFERENCES

BRAITHWAITE, R. B. (1953). *Scientific Explanation.* London.

EKMAN, G. (1958). Two generalized ratio scaling methods. *J. Psychol.*, **45**, 287–95.

LUCE, R. D., and GALANTER, E. (1963). Psychophysical scaling. In LUCE, R. D., BUSH, R. R., and GALANTER, E. (eds.). *Handbook of Mathematical Psychology.* Vol. I. New York and London, pp. 245–307.

PHILLIPS, J. P. N. (1964). A note on Treisman's model. *Quart. J. exp. Psychol.*, **16**, 386.

STEVENS, S. S. (1951). Mathematics, measurement and psychophysics. In STEVENS, S. S. (ed.). *Handbook of Experimental Psychology.* New York, pp. 1–49.

STEVENS, S. S. (1960). The psychophysics of sensory function. *Amer. Scientist*, **48**, 226–53.

STEVENS, S. S. (1964). Concerning the psychophysical power law. *Quart. J. exp. Psychol.*, **16**, 383–5.

TREISMAN, M. (1961). The nature of the psychophysical law and its significance for the scaling of sensory magnitudes. *Acta Psychol.*, **19**, 213–4.

TREISMAN, M. (1962). Psychological explanation: the 'private data' hypothesis. *Brit. J. phil. Sci.*, **13**, 130–43.

TREISMAN, M. (1963*a*). Laws of sensory magnitude. *Nature*, **198**, 914–5.

TREISMAN, M. (1963*b*). Neural effects and the psychophysical law. *Nature*, **200**, 1241–2.

TREISMAN, M. (1964*a*). Sensory scaling and the psychophysical law. *Quart. J. exp. Psychol.*, **16**, 11–22.

TREISMAN, M. (1964*b*). Noise and Weber's law: the discrimination of brightness and other dimensions. *Psychol. Rev.*, **71**, 314–30.

TREISMAN, M. (1964*c*). Signal detection theory and Crozier's Law: derivation of a new sensory scaling procedure. In preparation.

VOLKMANN, F. C. (1962). Vision during voluntary saccadic eye movements. *J. opt. soc. Amer.*, **52**, 571–8.

WARREN, R. M. (1958). A basis for judgments of sensory intensity. *Amer. J. Psychol.*, **71**, 675–87.

WHITE, C. T. (1963). Temporal numerosity and the psychological unit of duration. *Psychol. Monogr.*, **77**, Whole no. 575.

*Manuscript received 25th August, 1964.*

# BOOK REVIEWS

*British Medical Bulletin: Experimental Psychology*. Edited by A. Summerfield. London: Medical Department of the British Council. 1964. Vol. 20, Number 1. Pp. 86. 30s.

Medical research is increasingly relying on the methodology of the fundamental sciences, and the doctor who lacks basic training is at a disadvantage when it comes to the enlargement of medical knowledge. Among the sciences, a thorough grounding in psychology could prove rewarding, and it is towards the doctor who has already made himself familiar with basic psychological concepts that the bulletin under review is directed. This reduces the potential medical readership of this issue to a handful, but at the same time makes the bulletin suitable for a larger audience among psychologists. It offers material of potential clinical relevance, the scope of which is limited only by the range of the reader's imagination. No less important is the fact that a number of problems left unresolved in these reviews might perhaps be solved by suitable experiments on clinical case material.

Topics of current interest are reviewed by 15 leading experimental psychologists, all but one British. This patriotic approach to the selection of contributors results in the omission of some important recent work (particularly in the fields of animal psychology and of experimental linguistics). Nevertheless what is presented is of the greatest interest.

Reviews of human learning (Kay) and short-term memory (Brown) are complemented by more personal and highly readable accounts of selective attention (Treisman), vigilance (Broadbent) and decision making (Audley). Gregory discusses some perceptual issues, and Crossman's informational approach to skills winds up the first half of the symposium. The second half touches on a variety of issues, including an outstanding review of visual discrimination in animals (Sutherland). Weiskrantz used to best advantage the scant space allotted to him for a review of neurological studies and animal behaviour. Zangwill's account of neuropsychology, the only one directly concerned with clinical material, highlights this field in its uneasy transition from the anecdotal to the experimental stage. There are also articles on ageing (Welford), sleep (Oswald), and a pedestrian account of perceptual isolation (Zubek). Finally, reviews of drug effects on man and animals (Summerfield and Steinberg respectively) provide useful references for pharmacologists, though they might have been more suitably placed in the forthcoming bulletin on neuropharmacology.

Perhaps helped by editorial rigour, the contributors succeed in presenting accounts which are both concise and readable, and in which complex issues are explained with deceptive ease. The general impression is that of a well chosen and skilfully produced compilation, giving a vivid picture of the wealth of interesting experimental work at present under way in British laboratories.

M. KINSBOURNE.

*Reading Disability: Progress and Research Needs in Dyslexia*. Edited by John Money. Baltimore: The Johns Hopkins Press (London: Oxford University Press). 1962. Pp. x + 222. 40s.

This book must be welcomed as one of the first for many years to devote attention to the psychological processes actually involved in reading, and to make some attempt to pick out those which appear to be particularly affected in cases of developmental dyslexia. It is also to be commended in that it makes clear from the start that there is undoubtedly a general condition, termed specific or developmental dyslexia, which must be differentiated from conditions of reading backwardness caused by environmental factors such as inadequate teaching, or by general maladjustment. Moreover, it is stressed, notably by Zangwill, that large-scale surveys, in which various signs and factors are correlated one at a time with reading achievement, will throw little light on this condition. It is notable chiefly as a total syndrome or group of signs occurring together, none of which is unique in itself, but which are related in one or more specific configurations. This is emphasized by Money and Benton. Unfortunately, investigation of the incidence and exact relationships of symptoms and possible causal factors has often been insufficiently precise as regards testing, measurement and calculation of distribution to determine the nature of the syndromes; and clinicians must to some extent be to blame for this.

As with so many compendia of conference papers, these are uneven in value. One or two consist mainly of vague talk about the ascertainment and treatment of dyslexia. Two papers deal wholly with acquired dyslexia in adults, the relationship of which to developmental dyslexia is somewhat remote, as the authors admit. However, there remain some extremely valuable papers, including an excellent introductory paper by Money, which mark a real advance at least in suggestions as to the nature and causes of developmental dyslexia.

The principal point of agreement between a number of authors is that it has not been established that simple visual and auditory perception are deficient in true dyslexics. Some backwardness in visual and auditory perception, and even in directional perception, may retard the early stages of reading, but, as Benton emphasizes, most children grow out of this. In the true dyslexic the analysis of complex visual forms may be poor (Birch) and/or the processing and remembering of complex auditory patterns (Hardy). Birch suggests also that the organization of sensory patterns may be defective; in some children, the visual-auditory, in others the visual-kinaesthetic. But so far he has no very conclusive evidence on this. Perhaps most important is the hypothesis of Money, based on suggestions made by Rabinovitch, that conceptual difficulties are involved in the use of associated forms and sounds as symbols to which conceptual language meanings must be attached. The child must learn to understand that every word concept is symbolized by its constant and unique sequence of letters, which themselves may vary in certain respects (for instance, in different scripts), but not in others (for instance, orientation). Directional confusion may affect the remembering of the correct order of letters in sequences, but Benton is of opinion that in dyslexics it is the conceptualization and verbalizing of left-right directions which are deficient rather than their immediate perception.

Directional confusion has often been associated with weak or inconsistent laterality, which frequently appears in developmental dyslexia. Zangwill notes that this condition, often attributed to failure in development of normal cerebral dominance, may also be associated with some other deficiency in maturation or with unusual vulnerability to the effects of stress. These cases frequently show defects in speech development, spatial perception and motor control. On the other hand, there are cases of dyslexia in which none of these signs is apparent; clearly these require further investigation.

There is also the possibility that some other neurological dysfunction may be involved, or even minimal brain injury. Prechtl describes an interesting group of children with non-specific lesions of the brain who are characterized by excessive motor incoordination, with choreiform twitchings of the head and limbs. The eye muscles are affected, which may well cause difficulties in eye movement and fixation in reading. He has not yet demonstrated that this condition is associated with dyslexia; but other investigators have noted the frequent restlessness and over-activity of dyslexic children.

It may be concluded therefore that we have as yet no clear insight into the condition termed developmental dyslexia. Even its characteristic syndromes remain to be defined. But this book is of real value in that it suggests several significant points of departure for future investigation.

M. D. VERNON.

*Mathematical Methods in Small Group Processes.* Edited by Joan H. Criswell, Herbert Solomon and Patrick Suppes. London: Oxford University Press (for Stanford University Press). 1962. Pp. viii + 361. 78s.

This book is a collection of the 22 papers presented at the Stanford University Symposium in June, 1961, bringing together a distinguished set of contributors sharing an interest in "the process analysis of interpersonal relations." Its purpose is to set forth current developments in research on small groups, featuring mathematical methods and the experimentation to which they give rise.

Anyone new to this field will certainly be impressed by the amount and diversity of the work being done, but will search in vain for any unifying theme. This is perhaps unavoidable in a subject that is developing so fast, and so we must rely for the time being on picking out the odd interesting result and thoughtful comment that lie embedded in these papers.

Kurt Back, for example, strikes a fundamental note in claiming that many current experiments and models use a mathematics more suitable for things than for people, and urges more work on the social-communications aspect of experiments before any elaborate mathematics is applied. Among the most promising results we must include the Lorge-Solomon model for the prediction of group problem-solving efficiency from

the efficiencies of the individual group members, and its application by Restle to a waiting-time model for deriving the number of stages in the problem's solution. Both models will need to take account of the social interactions present before they agree well with experimental results. Foa has an interesting application of facet theory to husband-wife pairs, picking out three facets and showing a circular structure of the eight relationships, considered as elementary components of behaviour. The role of individual members in determining group standards or decisions is discussed in various ways by Anderson, Cohen, Moore and Anderson, Smoke and Zajonc, and Roby, whilst the influence of one group member upon another is the main interest of both Coleman and Karlsson, the latter giving a mathematical characterisation of the reward or punishment potential of the stimulus. Of the model-makers, Atkinson introduces a "strength of conditioning" factor to adapt previous models which he found defective, but Estes prefers to keep to the learning axioms and adopts a "scanning model" where the subjects choose the response that maximises the expected reward. Bush contributes a very thoughtful paper reaching pessimistic conclusions on the future of stochastic models, and insisting on the development of new experimental techniques before model-building can have any chance of success.
P. R. FREEMAN.

*Mental Radio.* By Upton Sinclair, with an Introduction by William McDougall, a Preface by Albert Einstein, and a Report by Walter Franklin Prince. Springfield, Illinois: C. C. Thomas. Revised Second Printing, 1962. Pp. xiv + 237.

In the first part of this volume Upton Sinclair, the well-known novelist and writer on social problems, describes his own experiments in the attempted telepathic "transmission" of rough drawings. In most of these experiments he was himself the "agent," and in all of them his wife was the "percipient." The experimental procedure was in many instances as follows. Mr. Sinclair, alone in his study, would prepare a small batch of drawings, place them in envelopes or otherwise conceal them, and take them to his wife. Mrs. Sinclair, lying in her room in a self-induced hypnoidal state (of which she gives a full description), would place the still-concealed drawings one by one over her solar plexus, and attempt to reproduce them. Of 290 attempted reproductions Mr. Sinclair ranks 65 as "successes," 155 as "partial successes," and the rest as failures. About half of the pairs of drawings are shown in the book, and it certainly does not seem plausible to ascribe the coincidences to chance. One might perhaps argue that persons who live together for many years will acquire a fund of common associations capable of producing such results; but against this must be set the fact that Mrs. Sinclair was nearly as successful with the 38 drawings provided by her husband's secretary as with those which Mr. Sinclair drew himself. And in any case, as Einstein says in his Preface, "if somehow the facts here set forth rest not upon telepathy, but upon some unconscious hypnotic influence from person to person, this also would be of high psychological interest. In no case should the psychologically interested circles pass over this book heedlessly."

The second part of this volume consists of a lengthy and able evaluation of the Sinclair experiments by Walter Franklin Prince, who carried out a large number of control experiments to rebut the hypothesis of chance coincidence. Dr. Prince also describes a further, and moderately successful, series of 25 attempted transmissions of drawings, with Mrs. Sinclair as the "percipient," and William McDougall as the "agent."

Those interested in extrasensory perception may find the experimental methods described in this book a profitable escape from the aching tediousness of the customary card tests.
A. GAULD.

*Clinical Correlations of Experimental Hypnosis.* Compiled and edited by Milton V. Kline. Springfield, Illinois: C. C. Thomas. 1963. Pp. xv + 624. $15.75.

*Hypnotism: An Objective Study in Suggestibility.* By A. M. Weitzenhoffer. London and New York: John Wiley (Science Editions). 1963. Pp. xvi + 380. 20s.

The first volume brings together 34 selected papers dealing with various aspe[illegible] hypnosis, of which six are original and the remainder culled from various journals—mostly American and mostly fairly recent. While a certain number are concerned with therapeutic applications of hypnosis (one of which bears the splendid title: "On the Psyche and Warts"), most of the papers deal with the experimental analysis of hypnotic phenomena and their relationship to other and less dramatic psychological issues. In the opinion of this reviewer, this collection is even more uneven and lacking in unity than most of its kind. But it will no doubt be of value to specialists.

The second volume is a paperback reissue of a book first published in 1953. On the whole, it is a well-informed and sensible study, with some attempt at theoretical interpretation. Most of the experimental studies relevant to hypnosis up to the year of original publication find mention.

O. L. Zangwill.

*The Mentally Retarded Child.* Edited by A. R. Luria. Translated from the Russian by W. P. Robinson. Oxford: Pergamon Press. 1963. Pp. viii + 207. 50s.

This is a popular exposition for teachers, psychiatrists and psychologists of work on mental defectives carried out by Luria and his colleagues. An opening chapter describes the causes of scholastic backwardness and distinguishes mental defect and its accompanying cortical pathology from pupils who have been neglected, those suffering from feeble or asthenic nervous systems and those who are deaf. The mentally defective group is then sub-divided aetiologically. The psychological consequences of cortical pathology namely the inertness of formed reflexes and lack of capacity to synthesise or generalize are mentioned as are also a notable lack of attention and concentration in these patients. Various possible causes are hypothesized for the psychological subtypes of mental defect described, i.e. the excitable, inhibited or balanced groups. A third chapter deals with the E.E.G. in imbeciles, a fourth with orienting reflexes. The fifth examines peculiarities of higher nervous function such as inertness or rigidity, lack of capacity to generalize on a conceptual basis and poor association between verbal and motor skills. Conditioning is slow and transfer on the basis of verbal instruction is poor, just as orientation was shown to be deficient in the previous chapter. Throughout these chapters external inhibition is shown to be stronger in imbeciles than in normals but orientation weaker. Two further chapters show firstly deficiencies in relationship between verbal and motor behaviour and secondly various aspects of verbal association behaviour. This material is already well known from previous translations of Luria's work. The central points of these chapters are the noted inability of defectives to make semantic connections rather than phonic ones. A short final chapter summarises the work reported.

The approach made in this book is both interesting and provocative and the original and ingenious experiments reported are of the same kind as those in Luria's previous publications. As this is a more popular work however, fewer details are given and some of the clinical observations would not accord with English experience or practice. This would also be true of some of the E.E.G. findings. It is to be hoped that through closer contact between Soviet and English workers in this field some of these differences of opinion can be resolved. In the meantime this small book could serve as a very good student's introduction to the Soviet approach to the psychology of mental defect.

N. O'Connor.

*Determinants of Infant Behaviour II.* Edited by B. M. Foss. With a Foreword by John Bowlby. London: Methuen. 1963. Pp. xii + 248. 45s.

This book contains the Proceedings of the Second Tavistock Seminar on mother-infant interaction held under the auspices of the Ciba Foundation in September, 1961. It is essentially a continuation of the first Symposium held 2 years earlier, the proceedings of which were reviewed in this *Journal*, 1963, **15**, 221. The present volume is 50 pages shorter and three shillings dearer. The contributors comprise H. F. Harlow, Thelma Rowell, H. F. R. Prechtl, Mary Ainsworth, Peter Wolff, Harriet Rheingold, H. R. Schaffer, J. A. Ambrose, R. A. Hinde and J. D. Sutherland. Mr. Brian Foss has once again done a competent editorial job but it is a pity that he does not otherwise contribute.

O. L. Zangwill.

*Scientific Principles of Psychology.* By Donald J. Lewis. New Jersey and London: Prentice-Hall. 1963. Pp. xi + 580. 64s.

*logy—the Study of Behaviour.* By Paul Swartz. Princeton, New York, Toronto and London: van Nostrand. 1963. Pp. xix + 451. 54s. 6d.

The American market for introductory psychological textbooks must be very large, and one may speculate how authors go about getting their share. Judging by these two texts, the aim is to cover all contemporary American psychology while giving the impression of an integrated discipline, and at the same time to emphasize a viewpoint. Both the history of the subject and controversial points should be played down; and, most important, the book should be visually attractive.

Lewis starts with methodology and statistics, then introduces his personal viewpoint, which is that psychology should be based on Pavlov and, especially, Skinner. A large part of the book is devoted to this, and it is done with thoroughness and clarity. Later there are sections on physiology, biology, motivation, perception and so on, but they link so poorly with Skinnerian psychology that they almost have the look of appendices, and no theory is developed with completeness. The book has its points for those who want an introduction to methodology and neo-behaviourism. However, when controversial points are mooted the author tends to take a dogmatic stand. Sometimes the argument is careless—"Humans apparently do not even possess an instinct for self-survival, for people under certain circumstances can be taught to kill themselves"—and there are inaccuracies even in statistical formulae (at one stage the printer has run out of capital sigmas and replaced them with capital Es). The glossary is a disaster: "Abscissa. The vertical line that frames a graph. The values of the measuring instrument are plotted on the abscissa"; "Culture. Any behaviour learned by a member of society from someone else"; "Identification. Responding to common social stimuli as if the responder were a larger, more powerful person or institution"; "Typology. A doctrine maintaining that body-type and personality are related."

Swartz's viewpoint is biological and social, and his chapters follow a sequence common in introductory texts. The result seems better integrated and personal bias is less evident. The experimental illustrations are carefully chosen, some being unusual. The author is honest about his sources, and he is accurate. The book is written at an elementary level, yet it would give a commendable scholarly impression if it were not for the method of presentation. The first chapter begins "Into a small, glass-enclosed box in the psychology laboratory of a large mid-western university a pleasant-faced young man in a white coat places a sleek hungry rat (see Figure 1.1)." Elsewhere the book is spattered with poems (which would make a lively anthology on their own) and lurid case-histories.

Nevertheless Swartz's book is by far the better of the two. One can imagine that it will be a success in the liberal arts colleges; and it would make a splendid text for the expanding British universities if someone would provide an English translation.

B. M. Foss.

*Multivariate Procedures for the Behavioral Sciences.* By William W. Cooley and Paul R. Lohnes. London and New York: Wiley. 1963. Pp. x + 211. 51s.

"This book describes some of the more useful multivariate techniques and how to compute and report them." Computer programmes are given as flow charts, from which complete programmes can be compiled for any digital computer, and also in FORTRAN, directly applicable to most IBM computers, and to an increasing number of others.

Even though the book both serves as an introduction and also supplies ready-to-run programmes, it does need to be supplemented by further instruction both in multivariate analysis and computing. References are given for this. A basic knowledge of statistics, but of no other mathematics, is assumed. The jargon may worry some readers, especially in the U.K.

The techniques are clearly, though not fully, described, making clear the type of problem for which each was designed. Most workers in the less exact sciences must have met with such problems. The authors stress that "computers are *easy* to use once the 'rules of the game' at the local center are understood." Probably only when the laborious computations involved in multivariate techniques are thus made easy, will researchers be able to evaluate dispassionately the results of the calculations. Recipe books like this one are therefore welcome.

PAUL WHITTLE.

*Perception tachistoscopique et communication.* By Arthur Fauville. Louvain: Nauwelaerts. 1963. Pp. 56. 80 FB.

This monograph is a review of experiments carried out at the Laboratoire de psychologie expérimentale de langage of the University of Louvain applying the information theory measure of channel capacity to absolute judgments in the visual modality. Since the details of most of the original experiments are only available as unpublished reports of the University of Louvain the monograph provides a useful summary of the work.

The technique used is the identification of the position of one or more points in a matrix presented tachistoscopically, and the main variables examined are size of matrix, relative position of the points and duration of exposure.

In general the results confirm American findings that when one increases the number of independent dimensions of variation in the stimulus the channel capacity of the subject is increased, but falls short of the value one would predict on the assumption that each dimension acted independently.

This is described in terms of a difference between "perception analytique" and "perception globale."

An analysis of errors reveals that these are not random, but that they show definite preferences or biases which are similar in all subjects.

The author does not, however, discuss the important point raised by Ericksen and Hake when dealing with multidimensional stimulus variations. They found that when variations along different dimensions were perfectly correlated, so that the total number of categories to be discriminated remained the same as if only one dimension had been used, channel capacity was increased by almost the same amount as with uncorrelated variations where the total number of stimulus categories is much greater.

As Garner points out, when one uses several dimensions of variation it does not seem to matter whether the stimulus variations are correlated or not. What is important is that the perceptual dimensions themselves should be relatively independent.

ROY DAVIS.

*Measurement in Personality and Cognition.* Edited by Samuel Messick and John Ross. New York and London: Wiley. 1963. Pp. xi + 334. 45s.

This book "is based on a conference, held at Princeton, New Jersey, under the auspices of the Educational Testing Service." It is divided into five sections, such as "Questionnaire Measures of Personality Dimensions" and "Stylistic Consistencies in Cognition," each section being followed by a Commentary, expounding and criticizing the relevant chapters.

From the laborious and heavy reading of this highly professional book, three points emerge: (i) That the terminology in this field is now past a joke. It has become a tragedy. (ii) That the yes/no and true/false type of item favoured in many personality inventories is inadequate and may be misleading. (iii) That the gulf is ever increasing, between the "personological" theorists and the man in the street—in whom the theorists claim to be interested.

A. W. HEIM.

*Maternal Behaviour in Mammals.* Edited by H. L. Rheingold. London and New York: Wiley. 1964. Pp. viii + 349. 66s.

This book contains ten studies of maternal behaviour in mammals, each reflecting the author's particular angle—behavioural, genetic or evolutionary. Those on moose and elk (Altman), langur monkeys (Jay) and baboons (De Vore) are based on observations of free-living animals: the two primate papers especially, written by anthropologists, contain much new and important material. The Wisconsin data on laboratory reared rhesus monkeys, reviewed by Harlow, Harlow and Hansen, provides a contrast in method of approach and demonstrates the progress which can be made by careful experimental work in this field. Herscher, Richmond and Moore emphasize the strong individual-specific mother-infant bond which is established in sheep and goats during a sensitive period shortly after birth, and its consequences for the development of the young. In the chapter on the rabbit (Ross, Sawin, Zarrow and Denenberg) the emphasis is on the interaction of genotypic and experimental factors in the control of maternal behaviour, several domesticated strains being compared. The study of Peromyscus (King), based on both field and laboratory observations, also involves the comparison of genetically different types—in this case two sub-species. Rheingold concentrates on the mother's behaviour in a study of the first few months of life of several varieties of dog, stressing at the same time the ways in which the mother's behaviour is modified by the developing young. This developmental approach is seen especially in the chapters on cat and rat (Schneirla, Rosenblatt and Tobach; Rosenblatt and Lehrman), which will surely become landmarks in the study of mother-infant interaction. Starting from a sound observational knowledge, but using precise recording techniques in well formulated experiments, these authors tease out the ways in which each developmental stage arises from the mother-infant-environment interactions of the preceding ones, working simultaneously at several biological and psychological levels.

The book is well illustrated by figures and plates, though the latter are not of uniform clarity.

R. A. HINDE.

*The Selective Process in the Functional Visual Field.* By A. F. Sanders. Soesterberg, Netherlands: Institute for Perception RVO-TNO Monograph. 1963. Pp. vi + 190.

The author, a Dutch psychologist at the Institute of Perception in Soesterberg, sees the selective process as a component of "attention," the other component being "intensity." In his introduction he shows how neglect of one or the other aspect led to "attention" either being equated with "set" or being merely a part of sensation, in both cases losing its usefulness as a separate concept. More recent uses of the word have been more profitable and Sanders scrutinizes particularly Broadbent's Filter Theory. This leads to a consideration of switching time, grouping and discrete selective acts from which the author emerges to describe some 20 experiments involving the simultaneous or sequential presentation of visual signals separated in space. As the visual angle is increased, the reasoning goes, the subject will be less able to group the stimuli and will have to treat them sequentially; i.e. using two selective acts. In this way we can hope to learn more about attention.

While the justification for the experiments is occasionally difficult to follow and not completely convincing (e.g. how far is it reasonable to talk about eye-fixations in reading as separate "selective acts"—and so as indications of changes in attention—when information regarding the same word may be taken in peripherally in two successive fixations?), the experiments themselves are well designed and informative.

Sanders is able to distinguish experimentally three levels of the functional visual field; the stationary field, the eyefield and the headfield, the boundaries of which, 20–30° and 75–105° respectively, coincide with drops in performance in various reaction time experiments. In the stationary field only one selective act is necessary (subject to information limitations) compared with two selective acts in the headfield. In the eyefield, a second selective act is often necessary, but only to confirm hypotheses already formulated on the basis of information gained in peripheral vision during the first fixation. Where two responses are required to simultaneously presented signals, the time for the second is less than that for the first except in the headfield where the times are equal. The author relates these results to the differences between grouping and successive handling, discusses alternative strategies for handling the information in the light of the work on refractoriness, and then looks at the effects of instructions and training upon performance. Finally, with more complex tasks, Sanders reports that the stationary field-eyefield discontinuity in performance tends to disappear, and the eyefield-headfield transition point varies with the difficulty of the task.

The author has the irritating habit of trying to draw smooth and fairly meaningless curves through three non-smooth points, and on occasion relies overmuch on subjective ratings from highly sophisticated subjects, but in the main the experiments are well treated. Sanders does not strain his data by constructing an overcomplex theory but rather regards his monograph as a pointer, which seems just. JOHN MORTON.

*Clinical Psychology: An Introduction to Research and Practice.* By N. D. Sundberg and L. E. Tyler. London: Methuen. 1963. Pp. xviii + 564. 50s.

This is a very "Global" book, covering every possible aspect of its field and using as many words as it can. 76 pages are devoted to discussing what clinical psychologists do, 200 odd to psychological assessment, and the rest to psychotherapy. At the end it is not much clearer than it was at the beginning either what psychologists *do* do or what good they are. MOYRA WILLIAMS.

*Expression of the Emotions in Man.* Edited by Peter H. Knapp, M.D. New York: International Universities Press (London: Bailey Bros. & Swinfen.) 1964. Pp. 351. 68s.

This is the report of a Symposium held at a meeting of the American Association for the Advancement of Science in December, 1960. It contains short contributions from some 30 psychiatrists, psychologists and sociologists, some of them very well known indeed. Yet the impression one gains is that this Symposium, like so many of its kind, was essentially contrived and that few of the authors had anything very fresh to say or felt any great incentive to say it. Even by modern standards the price seems inordinately high. O. L. ZANGWILL.

# INDEX OF AUTHORS

Andrews, D. P. 104
Anthony, W. S. 378
Baddeley, A. D. 364
Bahrick, H. P., with Bahrick, P. O. 318
Bahrick, P. O., with Bahrick, H. P. 318
Bitterman, M. E. with Wilson, W. A., and Oscar, M. 163
Brebner, J. with Gordon, I. 56
Broadbent, D. E., with Gregory, M. 309 and 359
Brown, J. 123
Burgh, P. 257
Butler, J. 272
Callaway, E., with Stone, G. C. 47
Cheng, M-F. H., with Davidon, R. S. 277
Crossman, E. R. F. W. 181
Davidon, R. S., with Cheng, M-F. H. 277
Davis, R. 332
Duncan, C. P. 373
Du Preez, P. D. 147
Gordon, I., with Brebner, J. 56
Gregory, M., with Broadbent, D. E. 309 and 359
Gregory, R. L., with Ross, H. E. 116
Gregory, R. L., with Shopland, C. D. 66
Groman, W. 184
Halliday, A. M., with Mingay, R. 35
Harper, R., with Stevens, S. S. 204
Hermelin, B. 241
Horn, G., with Venables, P. H. 289
Howard, I. P., with Templeton, W. B. 193
Hughes, I. M. 177
Jeeves, M. A. 61
Jeeves, M. A., with Pickersgill, M. J. 90
Keay, C. S. L., with Rodda, M. 282
Kimura, D. 355
Landauer, A. A. 23
Leonard, J. A. 325
McKean, K. O., with Miller, G. A. 297
Mackintosh, N. J. 250
Michon, J. A. 232
Miller, G. A., with McKean, K. O. 297
Mingay, R., with Halliday, A. M. 35
Morton, J. 340
Mowbray, G. H. 216
Mrosovsky, N. 166
Oscar, M., with Wilson, W. A., and Bitterman, M. E. 163
Pickersgill, M. J., with Jeeves, M. A. 90
Phillips, J. P. N. 386
Postman, L., with Turnage, T. W., and Silverstein, A. 81

Rabbitt, P. M. A. 172
Rodda, M., with Keay, C. S. L. 282
Ross, H. E., with Gregory, R. L. 116
Sampson, H. 1
Samuel, W. M. S. 264
Sanders, A. F. 368
Shopland, C. D., with Gregory, R. L. 66
Silverstein, A., with Postman, L., and Turnage, T. W. 81
Staniland, A. 134
Stevens, S. S. 383
Stevens, S. S., with Harper, R. 204
Stone, G. C., with Callaway, E. 47
Sutherland, N. S. 268
Templeton, W. B., with Howard, I. P. 193
Treisman, M. 11 and 387
Turnage, T. W., with Postman, L., and Silverstein, A. 81
Venables, P. H., with Horn, G. 289
Warr, P. B. 224 and 361
Wason, P. C. 30
Wells, M. J. 156
Wilson, W. A., with Oscar, M., and Bitterman, M. E. 163